Inexact trajectory planning and inverse problems in the Hamilton–Pontryagin framework

Christopher L. Burnett[1], Darryl D. Holm[1], David M. Meier[1]

Abstract

We study a trajectory-planning problem whose solution path evolves by means of a Lie group action and passes near a designated set of target positions at particular times. This is a higher-order variational problem in optimal control, motivated by potential applications in computational anatomy and quantum control. Reduction by symmetry in such problems naturally summons methods from Lie group theory and Riemannian geometry. A geometrically illuminating form of the Euler–Lagrange equations is obtained from a higher-order Hamilton–Pontryagin variational formulation. In this context, the previously known node equations are recovered with a new interpretation as Legendre–Ostrogradsky momenta possessing certain conservation properties. Three example applications are discussed as well as a numerical integration scheme that follows naturally from the Hamilton–Pontryagin principle and preserves the geometric properties of the continuous-time solution.

Contents

1 Introduction

The purpose of this paper. This paper is concerned with the analysis of a class of higher-order trajectory planning problems that are important in a wide range of contexts, ranging from computational anatomy to quantum control, both of which are discussed in this paper. The problem in its abstract formulation consists of finding an optimal curve $g(t)$ in a Lie group G that generates a curve $q(t) = g(t)Q_0$ in an object manifold Q, such that $q(t)$ passes near a set of fixed *target points* at given *node times*. Precise definitions of optimality and proximity will be provided below.

In the higher-order variational formulation appropriate for such problems, the Euler–Lagrange equations split into a set of Euler–Poincaré equations that hold on the open time intervals between the nodes, and a set of *node equations* that describe how to pass from one open interval to the next. The primary purpose of this paper is to develop a new geometric

understanding of the node equations in terms of conjugate momenta. Continuing from this, we develop a numerical algorithm for the trajectory planning problem that respects the geometric properties exhibited by the continuous-time solution.

1.1   Background and problem formulation

The problem treated in this paper fits into a classical type of problem in control theory called *trajectory planning*, or *interpolation by variational curves*. The task in this type of problem is to find an optimal curve that interpolates through a given set of points (or configurations) lying in some manifold, specific to the application one has in mind. Such trajectory planning problems are relevant in numerous applications, for example, in aeronautics, robotics, computer-aided design, air traffic control and more recently, computational anatomy. Some trajectory planning applications require the optimal trajectories to possess a certain degree of smoothness. This requirement summons variational principles that depend on higher-order derivatives of the interpolation path, such as *acceleration* (rate of change of velocity) or *jerk* (rate of change of acceleration) etc. Properties of such higher-order variational principles have been widely studied, one of the earlier references being [dLR85], where an intrinsic formulation in terms of higher-order tangent bundles was given. [PMRR11] contains a literature overview concerned both with mechanics and field theories, and some further recent developments are given in [GBHM+12, GBHR11, CdD11].

In some applications, for example in computational anatomy [GBHM+12] and quantum control [BHM12], the optimal trajectory is generated via a group action. That is, a curve $g(t)$ in a Lie group G acts on a point Q_0 in an *object manifold* and generates a curve $q(t) = g(t)Q_0$, that passes through a sequence of target points at prescribed times. In some instances it is desirable to relax the target constraints in such a way that the optimal curve does not exactly pass through the target points, but still passes *near* them at the prescribed times. This may be achieved by including a *soft constraint* in the cost functional, that is, a term penalizing the discrepancy between the trajectory $q(t)$ and the targets. This leads to cost functionals of the following type,

$$S = \int \ell(\xi, \dot{\xi}, \ldots, \xi^{(k-1)}) \, \mathrm{d}t + \frac{1}{2\sigma^2} \sum_i d^2(q(t_i), q_i). \tag{1.1}$$

Here, $\xi^{(j)}$ are the j-th time derivatives of a curve $\xi(t)$ in the Lie algebra (the tangent space at the identity element of the Lie group) that integrates to the curve $g(t)$, which in turn produces the trajectory in the object manifold according to $q(t) = g(t)Q_0$. The first part of the cost is the integral over a Lagrangian ℓ and is associated with the curve on the group. The second part sums up the squares of the distances d between the curve $q(t)$ and the target points q_i, at the prescribed times t_i. The *tolerance parameter* σ may be adjusted to suitably weight the two parts. In computational anatomy for example, a diffeomorphism group acts by transforming a medical image (or sub-structures thereof, such as points of interest, fibres, or surfaces). See [MTY02] for an overview. In this context the prescribed target images (or sub-structures) may not be diffeomorphically related to the initial one. That is, the initial and target configurations may lie on different group orbits. In general one expects therefore that exact matching may not be possible and works instead with a soft constraint.

Two of the authors have recently studied a problem of similar nature in quantum control, see [BHM12]. There one considers the group of unitary matrices acting on quantum state space, with the goal of finding the optimal experimental manipulation of the system such that the evolution of an initial state passes near a sequence of given target states. The cost functional is directly related to the required amount of change in the experimental apparatus over time. The introduction of a soft constraint is appropriate in this problem even though in this case the group action is transitive, since by increasing the tolerance parameter σ optimal trajectories may be achieved at a lower cost.

In a more general sense such trajectory planning problems can be thought of as *inverse problems*, where the data points $q_i \in Q$ have been determined experimentally at the times t_i and one seeks the corresponding curve $g(t)$ in configuration space G. In this context a natural choice of the tolerance parameter σ would be a measure of the uncertainty inherent in the experiment, such as standard deviation. The Lagrangian ℓ represents a modeling choice and is specific to the application one has in mind.

This paper is concerned with trajectory planning problems of the above type. It was found in [GBHM+12] that the Euler–Lagrange equations split into higher-order Euler–Poincaré equations, which hold on the open intervals between node times, and a set of *node equations* that describe the passage from one open interval to the next. These node equations describe the continuity properties of a set of quantities involving derivatives of various orders of the Lagrangian. As we will see in examples, a natural choice of Lagrangian ℓ leads to *Riemannian cubics* and their higher-order generalizations. This class of curves was introduced in [NHP89] and has since been studied in a series of papers including [CS95, CSC95, CSC01, Noa04, Noa06, GGP02]. Riemannian cubics appear in a variety of applications, for example in the quantum control problem mentioned above, but also in computer graphics, robotics and spacecraft control [PR97, ZKC98, HB04, VT12].

Plan of the paper.   In Section 2 we shall rederive the Euler–Lagrange equations for the higher-order variational problem by using Lagrange multipliers in a generalization of the symmetry reduced Hamilton–Pontryagin principle of geometric mechanics. In this approach, the derivation of the Euler–Lagrange equations simplifies considerably and a new geometric interpretation of the node equations emerges. Namely, they describe the evolution of Legendre–Ostrogradsky momenta across the nodes, in which the highest-order momentum experiences a discontinuous jump related to the distance between the curve in the object manifold and the target points. The discontinuity can be understood in terms of a momentarily broken symmetry at the node times. However, if the object manifold is isotropic with respect to a subgroup action then a residual symmetry remains. By Noether's theorem, this residual symmetry leads to a conservation law across node times.

In Section 3 we discuss a number of applications, including rigid body splines, macro-molecular configurations and quantum splines. Section 4 is concerned with the numerical solution of the inexact trajectory planning problem. More precisely, Section 4 describes a geometric discretization of the higher-order Hamilton–Pontryagin principle for inexact trajectory planning, similar to the approach given in [BRM09] for first-order systems. Our main motivation for the development of a geometric integrator is the exact momentum behavior of the discrete solution. This leads in turn to a dimensionality reduction of the search space

in the numerical optimization.

2   Geometry of the trajectory planning problem

We start with the statement of the problem considered here. One aims at steering from an initial point Q_0 in some *object manifold* Q along an optimal trajectory $q(t)$ that evolves via the action of a Lie group G. That is, $q(t) = g(t)Q_0$, where the right hand side denotes the action of $g(t)$ on Q_0 and the curve $q(t)$ lies in the G-orbit of Q_0.

The optimality condition is given in terms of a function $\ell : k\mathfrak{g} \to \mathbb{R}$ defined on the k-fold Cartesian product $k\mathfrak{g}$ of the Lie algebra $\mathfrak{g}$, which measures the cost of the transformation $g(t)$, and a distance function $d : Q \times Q \to \mathbb{R}$. As we shall see, the integer k determines the degree of smoothness of solution curves. The optimal curve $q(t)$ is required to pass near prescribed target points Q_{t_i} at prescribed *node times* t_i for $i = 1, \ldots, l$. This is formalized by including a squared distance term $d^2(g(t_i)Q_0, Q_{t_i})$ in the cost functional, for each i. Thus, the cost functional $S : \mathcal{C}(\mathfrak{g}) \to \mathbb{R}$, where $\mathcal{C}(\mathfrak{g})$ a suitable space of $\mathfrak{g}$-valued curves (see below) is defined by

$$S[\xi] := \int_0^{t_l} \ell(\xi, \ldots, \xi^{(k-1)}) \, \mathrm{d}t + \frac{1}{2\sigma^2} \sum_{i=1}^{l} d^2(g(t_i)Q_0, Q_{t_i}). \tag{2.1}$$

The notation $\xi^{(j)}$ is shorthand for $d^j\xi/dt^j$, the quantity σ is a *tolerance parameter*, and the curve $g(t)$ originates at the identity $g(0) = e$ and satisfies $\dot{g} = \frac{d}{d\varepsilon}\big|_{\varepsilon=0} \exp(\varepsilon\xi)g$. We write R_g for multiplication by g from the right and TR_g for its differential, thus $\dot{g} = TR_g\xi$. Variations are considered in the space $\mathcal{C}(\mathfrak{g})$ consisting of curves $\xi(t) : [0, t_l] \to \mathfrak{g}$ whose restrictions to open intervals (t_{i-1}, t_i) for $i = 1, \ldots, l$ are C^{2k-2} and whose j-th derivatives $\xi^{(j)}$ are continuous on $[0, t_l]$ for $j = 0, \ldots, k-2$. We also assume that initial values of $\xi^{(j)}(0)$ for $j = 0, \ldots, k-2$, are given.

This type of trajectory planning problem is familiar, for example, from image registration in computational anatomy, where one typically thinks of Q_0 as a template shape being deformed by a curve of diffeomorphisms $g(t)$, in turn generated by the time-dependent vector field $\xi(t)$ [You10]. At times t_i the resulting curve in shape space passes near the given target shapes Q_{t_i}, the parameter σ determining the proximity of the passage. In this case, the Lie group G of diffeomorphisms is infinite dimensional. However, in the present paper we will restrict ourselves to the case of finite-dimensional Lie groups and object manifolds. A natural finite-dimensional instance for illustrating these ideas arises in quantum control [BHM12], where quantum state vectors evolve under the action of the unitary group. The generator curve $\xi(t)$ in this case corresponds to the Hamiltonian operator, which is controlled in experiments.

2.1   Euler–Lagrange equations via Lagrange multipliers

The Euler–Lagrange equations characterize solutions to Hamilton's principle $\delta S = 0$ and were derived in [GBHM$^+$12]. The equations split into a set of higher-order Euler–Poincaré

equations on the open time intervals between the node times and a number of *node equations* describing how the solution evolves across the nodes. In previous formulations of this problem, one must find the variation $\delta g(t_i)$ that is produced by a (time dependent) variation $\delta \xi(t)$. This can be facilitated by taking advantage of Lagrange multipliers in an equivalent variational formulation that we describe now. As this paper demonstrates, the new approach also provides a geometric interpretation of the node equations and furthermore suggests a geometric numerical procedure for the solution of the problem , see Section 4 below.

The method of Lagrange multipliers involves enlarging the space on which the dynamics happen. We define the cost functional S on some space of curves $\mathcal{C}(G \times k\mathfrak{g} \times k\mathfrak{g}^*)$,

$$
\begin{aligned}
S[g, \xi^0, \dots, \xi^{k-1}, \mu^0, \dots, \mu^{k-1}] := \int_0^{t_l} &\left[\ell(\xi^0, \xi^1, \dots, \xi^{k-1}) + \langle \mu^0, TR_{g^{-1}}\dot{g} - \xi^0 \rangle \right. \\
&\left. + \sum_{r=1}^{k-1} \langle \mu^r, \dot{\xi}^{r-1} - \xi^r \rangle \right] dt + \frac{1}{2\sigma^2} \sum_{i=1}^{l} d^2(g(t_i)Q_0, Q_{t_i}).
\end{aligned}
\tag{2.2}
$$

Again some technical assumptions about the space of curves are needed. Namely the curves are C^1 when restricted to the open intervals (t_{i-1}, t_i) for $i = 1, \dots, l$, and $g, \xi^0, \dots, \xi^{k-2}$ are continuous on $[0, t_l]$. We also assume $g(0) = e$ and given initial values $\xi^j(0) = \xi_0^j$, for $j = 0, \dots, k-2$.

Before we take variations of S it is useful to introduce the *cotangent lift momentum map* $J^Q : T^*Q \to \mathfrak{g}^*$ associated with the action of G on Q (see [MR03], Chapter 11, for more information). This map is defined to satisfy

$$
\langle \alpha_q, \xi_Q(q) \rangle = \langle J^Q(\alpha_q), \xi \rangle, \quad \text{for any } \alpha_q \in T^*Q, \, \xi \in \mathfrak{g},
\tag{2.3}
$$

where we have used the notation $\xi_Q(q) := \frac{d}{d\varepsilon}\big|_{\varepsilon=0} e^{\varepsilon\xi} q$ and $\langle .,. \rangle$ for the respective duality pairings. We also introduce the shorthand $d_1 d(q_1, q_2) \in T_{q_1}^* Q$ to denote the exterior derivative of the distance function d with respect to the first entry, and $d_1 d(t_i) := d_1 d(g(t_i)Q_0, Q_{t_i})$. Integrating by parts and using (2.3) we obtain

$$
\begin{aligned}
\delta S = \int_0^{t_l} &\left[\langle -\dot{\mu}^0 - \text{ad}_{\xi^0}^* \mu^0, \eta \rangle + \sum_{r=0}^{k-2} \left\langle \frac{\delta\ell}{\delta\xi^r} - \mu^r - \dot{\mu}^{r+1}, \delta\xi^r \right\rangle + \left\langle \frac{\delta\ell}{\delta\xi^{k-1}} - \mu^{k-1}, \delta\xi^{k-1} \right\rangle \right. \\
&\left. + \langle \delta\mu^0, TR_{g^{-1}}\dot{g} - \xi^0 \rangle + \sum_{r=1}^{k-1} \left\langle \delta\mu^r, \dot{\xi}^{r-1} - \xi^r \right\rangle \right] dt \\
&+ \left\langle \mu^0(t_l) + \frac{d(t_l)}{\sigma^2} J^Q(d_1 d(t_l)), \eta(t_l) \right\rangle + \sum_{r=0}^{k-2} \left\langle \mu^{r+1}(t_l), \delta\xi^r(t_l) \right\rangle \\
&+ \sum_{s=1}^{l-1} \left[\left\langle \mu^0(t_s^-) - \mu^0(t_s^+) + \frac{d(t_s)}{\sigma^2} J^Q(d_1 d(t_s), \eta(t_s) \right\rangle + \sum_{r=0}^{k-2} \left\langle \mu^{r+1}(t_s^-) - \mu^{r+1}(t_s^+), \delta\xi^r(t_s) \right\rangle \right] \\
&- \left\langle \mu^0(0) - \frac{d(0)}{\sigma^2} J^Q(d_1 d(0)), \eta(0) \right\rangle - \sum_{r=0}^{k-2} \left\langle \mu^{r+1}(0), \delta\xi^r(0) \right\rangle,
\end{aligned}
\tag{2.4}
$$

where we set $\eta := TR_{g^{-1}}\delta g$ and defined $\mu^r(t_s^-) := \lim_{t \uparrow t_s} \mu^r(t)$ as well as $\mu^r(t_s^+) := \lim_{t \downarrow t_s} \mu^r(t)$. The last line above could have been omitted since by assumption $\eta(0) = 0$ and $\delta\xi^j = 0$ for $j = 0, \ldots, k-2$.

We can now read off the Euler–Lagrange equations. On the one hand, for t in any of the open intervals (t_i, t_{i+1}), $i = 0, \ldots, l-1$, we have

$$\dot{\mu}^0 + \mathrm{ad}^*_{\xi^0} \mu^0 = 0, \tag{2.5}$$

$$TR_{g^{-1}}\dot{g} - \xi^0 = 0, \tag{2.6}$$

$$\dot{\xi}^{r-1} - \xi^r = 0, \qquad (r = 1, \ldots, k-1) \tag{2.7}$$

$$\dot{\mu}^r + \mu^{r-1} - \frac{\delta\ell}{\delta\xi^{r-1}} = 0, \qquad (r = 1, \ldots, k-1) \tag{2.8}$$

$$\mu^{k-1} - \frac{\delta\ell}{\delta\xi^{k-1}} = 0. \tag{2.9}$$

On the other hand, the *node equations* are given by

$$\mu^0(t_s^-) - \mu^0(t_s^+) + \frac{d(t_s)}{\sigma^2} J^Q(\mathrm{d}_1 d(t_s)) = 0, \qquad (s = 1, \ldots, l-1) \tag{2.10}$$

$$\mu^r(t_s^-) - \mu^r(t_s^+), \qquad (r = 1, \ldots, k-1; \ s = 1, \ldots, l-1) \tag{2.11}$$

$$\mu^0(t_l) + \frac{d(t_l)}{\sigma^2} J^Q(\mathrm{d}_1 d(t_l)) = 0, \tag{2.12}$$

$$\mu^r(t_l) = 0. \qquad (r = 1, \ldots, k-1) \tag{2.13}$$

Remark 2.1. *There are 4 versions of the action functional, which are all relevant in applications. The one above can be called the* left-action, right-reduction *version since $g(t)$ acts on Q_0 from the left, while ξ^0 is the right-reduced velocity $\xi^0 = TR_{g^{-1}}\dot{g}$ (see Section 2.2 below, for more details on reduced variables). There are the following three other cases.*

(1) The right-action, right-reduction *case with action functional*

$$S[g, \xi^0, \ldots, \xi^{k-1}, \mu^0, \ldots, \mu^{k-1}] := \int_0^{t_l} \left[\ell(\xi^0, \xi^1, \ldots, \xi^{k-1}) + \langle \mu^0, TR_{g^{-1}}\dot{g} - \xi^0 \rangle \right.$$

$$\left. + \sum_{r=1}^{k-1} \langle \mu^r, \dot{\xi}^{r-1} - \xi^r \rangle \right] \mathrm{d}t + \frac{1}{2\sigma^2} \sum_{i=1}^{l} d^2(g(t_i)^{-1}Q_0, Q_{t_i}).$$

The variation of the penalty term in this case changes according to

$$\frac{1}{2}\delta d^2(g(t_i)^{-1}Q_0, Q_{t_i}) = -d(t_i) \left\langle \mathrm{Ad}^*_{g(t_i)^{-1}} J(\mathrm{d}_1 d(t_i)), \eta(t_i) \right\rangle.$$

This means that (2.10) and (2.12) are replaced by

$$\mu^0(t_s^-) - \mu^0(t_s^+) - \frac{d(t_s)}{\sigma^2} \mathrm{Ad}^*_{g(t_s)^{-1}} J^Q(\mathrm{d}_1 d(t_s)) = 0, \qquad (s = 1, \ldots, l-1)$$

$$\mu^0(t_l) - \frac{d(t_l)}{\sigma^2} \mathrm{Ad}^*_{g(t_l)^{-1}} J^Q(\mathrm{d}_1 d(t_l)) = 0.$$

(2) In the right-action, left-reduction *case, the action functional is*

$$S[G, \Xi^0, \ldots, \Xi^{k-1}, m^0, \ldots, m^{k-1}] := \int_0^{t_l} \left[l(\Xi^0, \Xi^1, \ldots, \Xi^{k-1}) + \langle m^0, TL_{G^{-1}}\dot{G} - \Xi^0 \rangle \right.$$

$$\left. + \sum_{r=1}^{k-1} \langle m^r, \dot{\Xi}^{r-1} - \Xi^r \rangle \right] dt + \frac{1}{2\sigma^2} \sum_{i=1}^{l} d^2(G(t_i)^{-1}Q_0, Q_{t_i}),$$

where we wrote L_G for multiplication by G from the left and TL_G for its differential. However, this is equivalent to the left-action, right-reduction case by identifying

$$G = g^{-1}, \quad \Xi^0 = -\xi^0, \ldots, \Xi^{k-1} = -\xi^{k-1}, \quad m^0 = -\mu^0, \ldots, m^{k-1} = -\mu^{k-1} \qquad (2.14)$$

and setting $\ell = l \circ \kappa$, where $\kappa : k\mathfrak{g} \to k\mathfrak{g}$ is multiplication by -1.

(3) By the same token the left-action, left-reduction *case with action functional*

$$S[G, \Xi^0, \ldots, \Xi^{k-1}, m^0, \ldots, m^{k-1}] := \int_0^{t_l} \left[l(\Xi^0, \Xi^1, \ldots, \Xi^{k-1}) + \langle m^0, TL_{G^{-1}}\dot{G} - \Xi^0 \rangle \right.$$

$$\left. + \sum_{r=1}^{k-1} \langle m^r, \dot{\Xi}^{r-1} - \Xi^r \rangle \right] dt + \frac{1}{2\sigma^2} \sum_{i=1}^{l} d^2(G(t_i)Q_0, Q_{t_i}).$$

can be mapped to the right-action, right-reduction case.

In the analysis that follows we will largely restrict ourselves to the left-action, right-reduction case. Anything we say can be transferred to the remaining three cases by applying the modifications listed above.

2.2 Euler–Poincaré equations

From (2.5)–(2.9) it follows that on open intervals (t_i, t_{i+1}), $i = 0, \ldots, l-1$,

$$\left(\frac{d}{dt} + \mathrm{ad}^*_{\xi^0} \right) \sum_{j=0}^{k-1} (-1)^j \frac{d^j}{dt^j} \frac{\delta\ell}{\delta\xi^j} = 0. \qquad (2.15)$$

This is a *k-th order Euler–Poincaré equation* for a system that exhibits right-invariance. This type of equation was derived in [GBHM$^+$12] from a variational perspective and in [CdD11] from the Hamiltonian one. We now explain its appearance in the inexact trajectory planning problem from the viewpoint of invariant Lagrangians, starting with some necessary definitions that can be found in [CMR01] and [GBHR11].

The *k-th order tangent bundle* $\tau_G^{(k)} : T^{(k)}G \to G$ is defined as a set of equivalence classes of curves as follows: Two curves $g_i(t) \in G, i = 1, 2$, are *equivalent*, if and only if their time derivatives at $t = 0$ up to order k coincide in any local chart. That is, $g_1^{(j)}(0) = g_2^{(j)}(0)$, for $0 \leq j \leq k$. The equivalence class of a curve $g(t)$ is denoted by $[g]_{g(0)}^{(k)}$, or formally as $(g(0), \dot{g}(0), \ldots, g^{(k)}(0))$. The set of all equivalence classes of curves emanating from $g_0 \in G$ is

written as $T_{g_0}^{(k)}G$. Then $T^{(k)}G := \bigcup_{g \in G} T_g^{(k)}G$ is a fibre bundle over G with projection map $\tau_G^{(k)} : [g]_{g(0)}^{(k)} \mapsto g(0)$. Note that a curve $g(t)$ defines, at each time t in its domain, an element $[g]_{g(t)}^{(k)} := [h]_{h(0)}^{(k)}$ by setting $h(s) := g(t + s)$.

It is convenient to represent $T^{(k)}G$ using the *trivialization map* that makes use of the right group multiplication (analogous constructions exist using the left multiplication map). Let $g(t)$ be a representative of $[g]_{g(0)}^{(k)}$ and define $\xi(t) := TR_{g^{-1}}\dot{g}(t)$. The trivialization map $T^{(k)}G \to G \times k\mathfrak{g}$ is given by

$$\alpha_k : [g]_{g(0)}^{(k)} \mapsto \big(g(0), \xi(0), \dot{\xi}(0), \dots, \xi^{(k-1)}(0)\big).$$

The *reduction map* $\beta_k : T^{(k)}G \to k\mathfrak{g}$ is obtained by omitting the first entry.

It is a well known (see, for example [MR03], Chapter 13) that the first order $(k = 1)$ Euler–Poincaré equation appears when the Euler–Lagrange equation for a Lagrangian $TG \to \mathbb{R}$ with symmetry is written in terms of the reduced velocity vector $\xi(t) = TR_{g^{-1}(t)}\dot{g}(t)$. The higher-order Euler–Poincaré equation appears in a similar fashion when computing the Euler–Lagrange equations for an invariant k-th order Lagrangian $L : T^{(k)}G \to \mathbb{R}$. More precisely, L is called *right-invariant* if $L\big([g]_{g(0)}^{(k)}\big) = L\big([gh]_{g(0)h}^{(k)}\big)$. The definition of *left-invariance* follows analogously. Let L be a right-invariant Lagrangian and consider Hamilton's principle, $\delta \mathcal{J} = 0$, for

$$\mathcal{J}[g] = \int_a^b L(g(t), \dot{g}(t), \dots, g^{(k)}(t)) \, \mathrm{d}t, \tag{2.16}$$

where variations are taken with respect to fixed end points up to order $k - 1$, that is, $\delta g^{(j)}(a) = \delta g^{(j)}(b) = 0$, for $j = 0, \dots k - 1$, in any local chart. The Euler–Lagrange equations can be written in terms of the right-reduced velocity vector, which leads to the *k-th order Euler–Poincaré equations* [GBHM+12]

$$\left(\frac{d}{dt} + \mathrm{ad}_\xi^*\right) \sum_{j=0}^{k-1} (-1)^j \frac{d^j}{dt^j} \frac{\delta \ell}{\delta \xi^{(j)}} = 0, \tag{2.17}$$

with *reduced Lagrangian* $\ell : k\mathfrak{g} \to \mathbb{R}$,

$$\ell\big(\xi(0), \dot{\xi}(0), \dots, \xi^{(k-1)}(0)\big) = L \circ \alpha_k^{-1}\big(e, \xi(0), \dot{\xi}(0), \dots, \xi^{(k-1)}(0)\big).$$

It is now straightforward to see from the viewpoint of invariant Lagrangians why the Euler–Poincaré equation (2.15) must characterize optimal curves in the trajectory planning problem on open time intervals. Indeed, fix $0 \le i \le l$ and suppose $\xi(t)$ is a local extremum of (2.1) with integral curve $g(t)$. If $\xi|_{(t_i, t_{i+1})}$ is not a solution to the k-th order Euler–Poincaré equation one can find a variation $g_\varepsilon(t)$ keeping fixed $g(t_i)$ and $g(t_{i+1})$, such that $\delta \int_{t_i}^{t_{i+1}} L(g, \dot{g}, \dots, g^{(k)}) \, \mathrm{d}t \ne 0$ with Lagrangian $L := \ell \circ \beta_k$. By consequence, $\xi_\varepsilon(t) := TR_{g_\varepsilon^{-1}(t)}\dot{g}_\varepsilon(t)$ is a variation of ξ such that $\delta S \ne 0$ for S as in (2.1), a contradiction.

2.3   Geometry of multipliers

Furthermore we observe from (2.5)–(2.9) that on open intervals (t_i, t_{i+1}), $i = 0, \ldots, l-1$,

$$\mu^r = \sum_{j=0}^{k-r-1} (-1)^j \frac{d^j}{dt^j} \frac{\delta\ell}{\delta\xi^{r+j}} \qquad (r = 0, \ldots, k-1). \tag{2.18}$$

In order to discuss the geometric meaning of these identities, we recall from above that the trajectory planning problem (2.2) on open intervals reduces to a problem of the type $\delta\mathcal{J} = 0$ with $\mathcal{J}$ of the form (2.16). In particular, equations (2.5)–(2.9) and therefore (2.18) are obtained by taking suitably constrained variations of

$$\mathcal{J}[g, \xi^0, \ldots, \xi^{k-1}, \mu^0, \ldots, \mu^{k-1}] := \int_a^b \left[\ell(\xi^0, \xi^1, \ldots, \xi^{k-1}) + \langle \mu^0, TR_{g^{-1}}\dot{g} - \xi^0 \rangle \right.$$
$$\left. + \sum_{r=1}^{k-1} \langle \mu^r, \dot{\xi}^{r-1} - \xi^r \rangle \right] \mathrm{d}t \tag{2.19}$$

This variational principle is a higher-order generalization of the *reduced Hamilton–Pontryagin principle* of first order mechanics. In first order mechanics, this principle provides a unified treatment of the Lagrangian and Hamiltonian descriptions of invariant mechanical systems on Lie groups (see [YM06] for a detailed discussion[1]). In particular, the *Legendre transform* connecting the two descriptions is revealed by the variational calculus. This remains true for higher-order mechanics. Indeed, (2.18) can be recognized to be the reduced Legendre transform that appears in [GBHM+12, GBHR11]. While we found (2.18) from a variational approach, these references take as starting point [dLR85], where a coordinate free description of the higher-order Legendre transform on manifolds was given. We briefly review this approach here.

The Legendre transform of higher-order mechanics, given in [dLR85], is a map Leg : $T^{(2k-1)}G \to T^*(T^{(k-1)})G$. If Leg is a diffeomorphism (that is, L is *hyperregular*) it connects the Lagrangian and Hamiltonian descriptions just as in the first order case. With respect to the right-trivializations

$$T^{(2k-1)}G \cong G \times (2k-2)\mathfrak{g}, \qquad T^*(T^{(k-1)}) \cong G \times (k-2)\mathfrak{g} \times (k-1)\mathfrak{g}^* \tag{2.20}$$

it is given as [GBHR11]

$$\text{Leg}: (g, \xi^0, \ldots, \xi^{2k-2}) \mapsto (g, \xi^0, \ldots, \xi^{k-2}, \mu^0, \ldots, \mu^{k-1}),$$
$$\text{where} \quad \mu^r = \sum_{j=0}^{k-r-1} (-1)^j \frac{d^j}{dt^j} \frac{\delta\ell}{\delta\xi^{r+j}} \quad (r = 0, \ldots k-1).$$

The same equations were seen in (2.18) to emerge from the Hamilton–Pontryagin principle (2.19). This means that, as for first order, the higher-order Hamilton–Pontryagin principle

---

[1]There is a close connection between the Hamilton–Pontryagin principle and Dirac structures, an aspect we do not enter into in the present paper. See [YM06] for more information.

contains both Lagrangian and Hamiltonian descriptions of higher-order mechanics and provides a unified framework for both views. To obtain the Lagrangian description one may eliminate $\mu^0, \ldots, \mu^{k-1}$ from (2.5)–(2.9) using (2.18). The resulting equations are the trivialized flow equations of the *Lagrangian vector field*, an element of $\mathfrak{X}(T^{(2k-1)}G)$ [dLR85]. On the other hand, if (2.9) can be solved for ξ^{k-1} (this is the case, for example, when L is hyperregular) then (2.5)–(2.8) are the trivialized flow equations of the *Hamiltonian vector field* $X_H \in \mathfrak{X}(T^*(T^{(k-1)}))$, which solves [dLR85]

$$i_{X_H}\omega = \mathrm{d}H, \qquad (2.21)$$

where ω is the canonical symplectic form on $T^*(T^{(k-1)}G)$ and $H : T^*(T^{(k-1)}G) \to \mathbb{R}$ is given as

$$H(g, \xi^0, \ldots, \xi^{k-2}, \mu^0, \ldots \mu^{k-1}) = \sum_{r=0}^{k-1} \langle \mu^r, \xi^r \rangle - \ell(\xi^0, \ldots, \xi^{k-1}) \qquad (2.22)$$

with respect to the trivialization (2.20). By consequence of (2.21) the flow map $F_t : T^*(T^{(k-1)}) \to T^*(T^{(k-1)})$ of the Hamiltonian vector field preserves the symplectic form ω.

   For later reference we point out how this can be seen alternatively from the Hamilton–Pontryagin principle. This is a generalization to higher order of a standard argument (see for example [BRM09], Section 3, for the first order case). If we omit end point constraints on the variations of $\mathcal{J}$ in (2.19), the integration by parts contributes boundary terms to $\delta\mathcal{J}$ (cf. (2.4)),

$$\delta\mathcal{J} = \int_a^b \cdots \mathrm{d}t + \langle \mu^0, TR_{g^{-1}}\delta g\rangle\big|_a^b + \sum_{r=0}^{k-2} \langle \mu^{r+1}, \delta\xi^r\rangle\big|_a^b = \int_a^b \cdots \mathrm{d}t + \theta(\delta x)\big|_a^b, \qquad (2.23)$$

where θ in the second equality is the canonical one-form on $T^*(T^{(k-1)}G)$ and we defined $\delta x(t)$ to be the curve in $TT^*(T^{(k-1)}G)$ whose trivialization corresponds to the variations $(TR_{g^{-1}}\delta g, \delta\xi^0, \ldots, \delta\xi^{(k-2)}, \delta\mu^0, \ldots, \delta\mu^{(k-1)})$. If we restrict the variations to solution curves of (2.5)–(2.9), we may just as well express $\mathcal{J}$ as a function of initial conditions $\mathcal{J}_{\text{initial}} : T^*(T^{(k-1)}G) \to \mathbb{R}$. The integral part of (2.23) then vanishes and we obtain

$$\delta\mathcal{J} = \mathrm{d}\mathcal{J}_{\text{initial}}(\delta x(a)) = (F_{b-a}^*\theta - \theta)(\delta x(a)).$$

Therefore, $F_{b-a}^*\theta = \theta$, and taking an exterior derivative yields the desired identity $F_{b-a}^*\omega = \omega$.

2.4   Momentum conservation and Noether's theorem

Another important point to notice from (2.5)–(2.9) is that $\mathrm{Ad}_g^*\mu^0$ is a conserved quantity. Here Ad^* is the map dual to $\mathrm{Ad} : G \times \mathfrak{g} \to \mathfrak{g}$ given by $(g, \xi) \mapsto \mathrm{Ad}_g\xi := TL_gTR_{g^{-1}}\xi$. Indeed by (2.5)

$$\frac{d}{dt}\,\mathrm{Ad}_g^*\,\mu^0 = \mathrm{Ad}_g^*(\dot\mu^0 + \mathrm{ad}_{\xi^0}^*\,\mu^0) = 0. \qquad (2.24)$$

In the context of first order Euler–Poincaré equations a similar momentum conservation is due to the invariance of the Lagrangian with respect to group multiplication operations. This is an instance of Noether's theorem, which roughly speaking guarantees that the momentum map associated with the action of a symmetry group is preserved (see for example [MR03], Chapter 11). We now show that the situation is similar for (2.24). The right action R of G on itself,

$$R : G \times G, \quad (h, g) \mapsto R_g(h) = hg,$$

can be lifted to an action on $T^{(k-1)}G$,

$$T^{(k-1)}R : T^{(k-1)}G \times G \to T^{(k-1)}G, \quad \left([h]_{h(0)}^{(k-1)}, g\right) \mapsto T^{(k-1)}R_g\left([h]_{h(0)}^{(k-1)}\right) = [hg]_{h(0)g}^{(k-1)}.$$

This action can be lifted to its so-called *cotangent lifted action* ([MR03], Section 12.1)

$$T^*T^{(k-1)}R : T^*(T^{(k-1)}G) \times G \to T^*(T^{(k-1)}G),$$

given in trivialized form as

$$T^*T^{(k-1)}R_g(h, \xi^0, \dots \xi^{k-2}, \mu^0, \dots \mu^{k-1}) = (hg^{-1}, \xi^0, \dots, \xi^{k-2}, \mu^0, \dots \mu^{k-1}).$$

It is apparent that the Hamiltonian (2.22) is symmetric with respect to this group action. By Noether's theorem the associated momentum map is conserved.

What is this momentum map? By appealing to standard formulas ([MR03], Section 12.1) we find that the momentum map $J : T^*(T^{(k-1)}G) \to \mathfrak{g}^*$ is

$$J(g, \xi^0, \dots \xi^{k-2}, \mu^0, \dots \mu^{k-1}) = \mathrm{Ad}_g^* \mu^0,$$

with respect to the trivialization (2.20). The conservation law observed in (2.24) thus arises from the right-invariance of the Lagrangian (respectively, the Hamiltonian) via Noether's theorem.

The conservation law can also be obtained from a variational perspective. This is well known in first order mechanics, and it is also the case in higher-order mechanics. We take a solution of (2.5)–(2.9) on the time interval $[a, b]$ and vary it according to $\delta g = TL_g \nu$ for $\nu \in \mathfrak{g}$. For $\mathcal{J}$ as in (2.19) we have (cf. (2.23))

$$\delta \mathcal{J} = 0 = \left\langle \mu^0, TR_{g^{-1}} TL_g \nu \right\rangle \big|_a^b = \left\langle \mathrm{Ad}_g^* \mu^0, \nu \right\rangle \big|_a^b. \tag{2.25}$$

The same argument holds after replacing the upper boundary b by any $b' \in [a, b]$. Since ν was arbitrary we conclude that $\mathrm{Ad}_g^* \mu^0$ is conserved along a solution of (2.5)–(2.9).

2.5   Node equations

The remarks above concerned equations (2.5)–(2.9) on the open time intervals between nodes. We now come to the node equations (2.10)–(2.13). These specify the evolution across node times of the Lagrange multipliers μ^r, which we interpreted above as the reduced *Legendre*

momenta of the system. More specifically, the momenta μ^r, $r = 1, \ldots, k-1$ are continuous on $[0, t_l]$, while the 0-th momentum μ^0 experiences jump discontinuities at the nodes. If the Lagrangian ℓ is hyperregular we can conclude that $g \in C^{2k-2}([0, t_l])$, that is, g is $(2k-2)$ times continuously differentiable on $[0, t_l]$. Furthermore, the node equations specify terminal values for the curves μ^r, $r = 1, \ldots, k-1$.

For a, $b \in \mathbb{R}$ define $1_{a \leq b}$ to be equal to 1 if $a \leq b$ and 0 otherwise. We can now prove the following theorem.

Theorem 2.2. *For t in any of the open time intervals (t_s, t_{s+1}) as well as for $t \in \{0, t_l\}$,*

$$\mu^0(t) = -\frac{1}{\sigma^2} \mathrm{Ad}^*_{g(t)^{-1}} \left(\sum_{s=1}^{l} 1_{t \leq t_s} d(t_s) \, \mathrm{Ad}^*_{g_{t_s}} J^Q(\mathrm{d}_1 d(t_s)) \right). \tag{2.26}$$

Proof. At final time $t = t_l$ (2.26) clearly holds because of (2.12). Since $\mathrm{Ad}^*_g \mu^0$ is conserved on open intervals it follows that for $t \in (t_s, t_{s+1})$,

$$\mu^0(t) = \mathrm{Ad}^*_{g(t)^{-1}} \mathrm{Ad}^*_{g(t_{s+1})} \mu^0(t_{s+1}^-).$$

We can now obtain (2.26) by induction over the open time intervals, noting that at each node $t = t_s$ a term $-\frac{d(t_s)}{\sigma^2} J^Q(\mathrm{d}_1 d(t_s))$ gets added on. ∎

In order to formulate the following corollary we use the notation $\mathfrak{g}_q$ for any point $q \in Q$ to denote the Lie algebra of the *isotropy subgroup* of that point, $G_q := \{g \in G | gq = q\}$. In particular $\rho_Q(q) = 0$ for any $\rho \in \mathfrak{g}_q$.

Corollary 2.3. *For a solution of (2.5)–(2.13) we have, for t in any of the open time intervals (t_s, t_{s+1}) as well as for $t \in \{0, t_l\}$,*

$$\langle \mu^0(t), \rho \rangle = 0 \qquad \text{for all } \rho \in \mathfrak{g}_{q(t)}. \tag{2.27}$$

In Section 4 we will develop a geometric algorithm that inherits an exact version of this Corollary. This implies that the numerical search for the optimal initial value of μ^0 can be restricted to the subspace of $\mathfrak{g}^*$ that annihilates $\mathfrak{g}_{Q_0}$.

Proof. For t and ρ as in the statement of the corollary it follows from Theorem 2.2 that

$$\langle \mu^0(t), \rho \rangle = -\frac{1}{\sigma^2} \sum_{s=1}^{l} 1_{t \leq t_s} \left\langle d(t_s) J^Q(\mathrm{d}_1 d(t_s)), \mathrm{Ad}_{g_{t_s}} \mathrm{Ad}_{g(t)^{-1}} \rho \right\rangle$$

$$= -\frac{1}{\sigma^2} \sum_{s=1}^{l} 1_{t \leq t_s} \left\langle d(t_s) \, \mathrm{d}_1 d(t_s), \left(\mathrm{Ad}_{g_{t_s}} \mathrm{Ad}_{g(t)^{-1}} \rho \right)_Q (q(t_s)) \right\rangle = 0,$$

where we used (2.3) for the second equality and noted that $\mathrm{Ad}_{g_{t_s}} \mathrm{Ad}_{g(t)^{-1}} \rho \in \mathfrak{g}_{q(t_s)}$ for the third. ∎

2.6   Residual conservation law after partial symmetry breaking

A physically intuitive perspective on Corollary 2.3 is to understand it as a residual conservation law after partial symmetry breaking. We can see the sum in (2.2) as the integral over a time-dependent *potential function* $V : [0, t_l] \times G \to \mathbb{R}$ given by

$$V(t, g) = \frac{1}{2\sigma^2} \sum_{i=1}^{l} \delta(t - t_i) \, d^2(g(t)Q_0, Q_{t_i}) \tag{2.28}$$

This potential produces instantaneous singular forces at node times t_i which impart the jump discontinuities on the otherwise conserved momentum $J = \mathrm{Ad}_g^* \mu^0$,

$$J(t_s^+) = J(t_s^-) + \frac{d(t_s)}{\sigma^2} \mathrm{Ad}_{g(t_s)}^* J^Q(\mathrm{d}_1 d(t_s)). \tag{2.29}$$

This is because the presence of this potential breaks the G-invariance of the variational problem, however a residual symmetry remains. Clearly, multiplication of g from the right by an element $h \in G_{Q_0}$ leaves V invariant. An adaptation of the argument surrounding equation (2.25), restricting ν to the subspace $\mathfrak{g}_{Q_0} \subset \mathfrak{g}$, then leads to

$$0 = \left\langle \mathrm{Ad}_g^* \mu^0, \nu \right\rangle \big|_0^t .$$

for any $t \in [0, t_l]$. Moreover, (2.12) guarantees that $\left\langle \mathrm{Ad}_{g(t_l)}^* \mu^0(t_l), \nu \right\rangle = 0$. Therefore, $\left\langle \mathrm{Ad}_{g(t)}^* \mu^0(t), \nu \right\rangle = 0$ for any $t \in [0, t_l]$ and $\nu \in \mathfrak{g}_{Q_0}$, which is equivalent to Corollary 2.3.

3   Applications

In this section we discuss a number of examples that summon the inexact trajectory planning problem. We start with a discussion of Riemannian cubics on general Riemannian manifolds and specifically on Lie groups with invariant metrics, since this type of curve appears for a certain natural choice of Lagrangian. We then treat the rigid body, finite-dimensional quantum systems and molecular strands.

3.1   Riemannian cubics

Let (M, γ) be a Riemannian manifold with metric γ, whose norm we denote by $\|.\|_\gamma$. The notion of straight lines in Euclidean space generalizes to geodesics in M. These are curves $x(t) \in M$ that satisfy $D_t \dot{x} = 0$. Here we introduce the notation $D_t \dot{x} = \nabla_{\dot{x}} \dot{x}$, where ∇ is the Levi–Civita connection for γ. In local coordinates the geodesic equation is given by

$$\ddot{x}^k + \Gamma_{ij}^k \dot{x}^i \dot{x}^j = 0,$$

where Γ_{ij}^k are the Christoffel symbols of the Levi–Civita connection. We define the vector bundle isomorphism over the identity, $\flat : TM \to T^*M$, which maps $v \in T_x M$ to $v^\flat = \gamma(x)(v_x, \cdot)$. The inverse map is denoted by $\sharp$.

The physical interpretation of straight lines in Euclidean space follows from Newton's second law $\ddot{x} = F$. Here F is an external force acting on a particle of unit mass that follows the trajectory $x(t)$. By analogy one can often understand $(D_t \dot{x})^\flat$ as a generalized force acting on a physical system whose time evolution is represented by a curve $x(t)$ in configuration space M. Consider the Lagrangian $L : T^{(2)}M \to \mathbb{R}$,

$$L(x, \dot{x}, \ddot{x}) = \frac{1}{2} \left\| D_t \dot{x} \right\|_\gamma^2 , \tag{3.1}$$

whose corresponding action functional $S = \int_a^b L \, \mathrm{d}t$ measures the square of the L^2-norm of the external force. Hamilton's principle $\delta S = 0$ for variations with fixed boundary velocities $\dot{x}(a)$ and $\dot{x}(b)$ leads to the Euler–Lagrange equation [NHP89]

$$D_t^3 \dot{x} + R(D_t \dot{x}, \dot{x})\dot{x} = 0 \,,$$

where R is the curvature tensor, $R(X, Y)Z := \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z$ for any vector fields $X, Y, Z \in \mathfrak{X}(M)$. Solutions to this equation are called *Riemannian cubics* and were introduced in [NHP89].

When the manifold is a Lie group $M = G$ we call a Riemannian metric *right (or left) invariant* if $\gamma(g)(v_g, w_g) = \gamma(gh)(TR_h v_g, TR_h w_g)$, or $\gamma(g)(v_g, w_g) = \gamma(hg)(TL_h v_g, TL_h w_g)$ respectively, for all g, $h \in G$ and v_g, $w_g \in T_g G$. If the metric is right (or left) invariant, then the Lagrangian (3.1) can be reduced to a function $\ell : 2\mathfrak{g} \to \mathbb{R}$ [GBHM$^+$12]

$$\ell(\xi^0, \xi^1) = \frac{1}{2} \left\| \xi^1 \pm \mathrm{ad}_{\xi^0}^\dagger \xi^0 \right\|_\gamma^2 , \tag{3.2}$$

where we introduced the operation $\mathrm{ad}_\xi^\dagger \rho = (\mathrm{ad}_\xi^* \rho^\flat)^\sharp$. This can be seen from the fact (e.g., [GBHM$^+$12]) that for a curve $g(t)$, whose right (or left) reduced velocity is given by $\xi(t) \in \mathfrak{g}$,

$$D_t \dot{g} = \left(\dot{\xi} + \mathrm{ad}_\xi^\dagger \xi \right) g, \quad \text{or} \quad D_t \dot{g} = g \left(\dot{\xi} - \mathrm{ad}_\xi^\dagger \xi \right) .$$

In the following we will discuss a number of physical systems whose configuration spaces are Lie groups and whose equation of motion in the absence of external forces is given by $D_t \dot{g} = 0$ for some right (or left) invariant Riemannian metric. The Lagrangian (3.2) is then one natural choice for ℓ in the inexact trajectory planning problem since optimal curves minimize (in the L^2 sense) the amount of external forcing necessary to achieve them. It is clear from equations of motion (2.5)–(2.11) that the solution $g(t)$ is a Riemannian cubic on open intervals, and twice continuously differentiable on the whole time interval $[0, t_l]$. Such curves are called *Riemannian cubic splines*.

Remark 3.1. *Without going into the mathematical details we point to a probabilistic interpretation of Riemannian cubics. See also [VT12], where a closely related idea is discussed in the context of stochastic modeling of biological growth. Let G be a Lie group with right-invariant metric γ and let e_i, $i = 1, \ldots, d$ be an orthonormal basis of the d-dimensional Lie algebra $\mathfrak{g}$. Consider a curve $g(t) \in G$, whose right-reduced velocity $\xi = TR_{g^{-1}}\dot{g}$ satisfies the following Ito stochastic differential equation*

$$\mathrm{d}\xi = -\,\mathrm{ad}_\xi^\dagger \xi \, \mathrm{d}t + \sigma_W \sum_{i=1}^d \mathrm{d}W_i e_i, \tag{3.3}$$

where W_i, $i = 1, \ldots, d$, are independent Brownian motions (see, for example, [Øks03] Section 2.2, for a definition) and $\sigma_W \in \mathbb{R}$. Suppose the (noisy) data is given in a vector space V equipped with an inner product, whose norm we denote by $\|.\|_V$. The noise distribution is assumed to be Gaussian, that is, the probability density function has the form $p(Q) \sim \exp(-\frac{1}{2\sigma_n^2}\|Q - \bar{Q}\|_V^2)$, where $\bar{Q}$ is the true state of the system and $\sigma_n \in \mathbb{R}$. Suppose experiments at times t_i, $i = 1, \ldots, l$ measuring the trajectory $g(t)Q_0$ have given results Q_{t_i}. Then the minimization of

$$S = \int_0^{t_l} \ell(\xi, \dot{\xi}) \, \mathrm{d}t + \frac{\sigma_W^2}{2\sigma_n^2} \sum_{i=1}^{l} \|g(t_i)Q_0 - Q_{t_i}\|_V^2,$$

with ℓ as in (3.2), can formally be understood as the maximization of the (logarithm of the) probability of the path $g(t)$, given the measurements. Alternative models of stochastic forcing will typically lead to minimization problems of the same type, but with different choices of ℓ (see Remark 3.2 below).

3.2   Rigid body splines

Let the Lie group G be the set of rigid rotations $SO(3)$, and let Q be the unit sphere $S^2 \in \mathbb{R}^3$. We use vector notation for the Lie algebra $\mathfrak{so}(3)$ of the Lie group of rotations $SO(3)$, as well as for its dual $\mathfrak{so}(3)^*$. One identifies $\mathfrak{so}(3)$ with $\mathbb{R}^3$ via the familiar isomorphism

$$\widehat{} : \mathbb{R}^3 \to \mathfrak{so}(3), \quad \mathbf{\Omega} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \mapsto \Omega := \widehat{\mathbf{\Omega}} = \begin{pmatrix} 0 & -a & b \\ a & 0 & -c \\ -b & c & 0 \end{pmatrix}, \tag{3.4}$$

called the *hat map*. This is a Lie algebra isomorphism when the vector cross product $\times$ is used as the Lie bracket operation on $\mathbb{R}^3$. The identification of $\mathfrak{so}(3)$ with $\mathbb{R}^3$ induces an isomorphism of the dual spaces $\mathfrak{so}(3)^* \cong (\mathbb{R}^3)^* \cong \mathbb{R}^3$, with the dot product as duality pairing. Let γ be a left-invariant Riemannian metric on $SO(3)$. This defines an inner product on $\mathfrak{so}(3)$ which can be expressed as

$$\gamma_e(\mathbf{\Omega}_1, \mathbf{\Omega}_2) = \mathbf{\Omega}_1 \cdot \mathbb{I}\mathbf{\Omega}_2$$

for a symmetric, positive definite matrix $\mathbb{I}$. The geodesic equation $D_t \dot{g} = 0$ in Euler–Poincaré form is

$$\dot{\mathbf{\Omega}} + \mathbb{I}^{-1}(\mathbf{\Omega} \times \mathbb{I}\mathbf{\Omega}) = 0, \qquad \dot{g} = g\widehat{\mathbf{\Omega}}. \tag{3.5}$$

Consequently, the Lagrangian (3.2) takes the form

$$\ell(\mathbf{\Omega}^0, \mathbf{\Omega}^1) = \frac{1}{2}(\mathbf{\Omega}^1 + \mathbb{I}^{-1}(\mathbf{\Omega}^0 \times \mathbb{I}\mathbf{\Omega}^0)) \cdot \mathbb{I}(\mathbf{\Omega}^1 + \mathbb{I}^{-1}(\mathbf{\Omega}^0 \times \mathbb{I}\mathbf{\Omega}^0))$$

$$= \frac{1}{2} \left\| \mathbf{\Omega}^1 + \mathbb{I}^{-1}(\mathbf{\Omega}^0 \times \mathbb{I}\mathbf{\Omega}^0) \right\|_{\mathfrak{so}(3)}^2. \tag{3.6}$$

Consider the inexact trajectory planning problem in the left-action, left-reduction form. That is, suppose an initial point $\mathbf{x}_0$ and targets $\mathbf{x}_{t_1}, \ldots \mathbf{x}_{t_l} \in S^2$ are given, as well as a tolerance parameter σ. We seek the minimizer of

$$S[g, \mathbf{\Omega}^0, \mathbf{\Omega}^1, \boldsymbol{\mu}^0, \boldsymbol{\mu}^1] = \int_{t_0}^{t_l} \frac{1}{2} \left\| \mathbf{\Omega}^1 + \mathbb{I}^{-1}(\mathbf{\Omega}^0 \times \mathbb{I}\mathbf{\Omega}^0) \right\|_{\mathfrak{so}(3)}^2 + \left\langle \boldsymbol{\mu}^0, g^{-1}\dot{g} - \mathbf{\Omega}^0 \right\rangle$$

$$+ \left\langle \boldsymbol{\mu}^1, \dot{\mathbf{\Omega}}^0 - \mathbf{\Omega}^1 \right\rangle + \frac{1}{2\sigma^2} \sum_{i=1}^{l} \left\| g(t_i)\mathbf{x}_0 - \mathbf{x}_{t_i} \right\|^2. \tag{3.7}$$

The physical interpretation is as follows. The group of rigid rotations, $SO(3)$, is the configuration manifold of a rigid body constrained to rotate around a fixed point. In the absence of external torques the motion is governed by the geodesic equation (3.5), where $\mathbb{I}$ is the moment of inertia tensor (see, for example, [Hol11], Section 2.4). The resulting curve $g(t)$ describes the orientation of the rigid body relative to a space-fixed reference frame.

Suppose the motion of a rigid body (with or without external torque) is partially observed in an experiment. At discrete times t_i the direction of a particular body fixed axis is measured in the space-fixed frame, generating a sequence of outcomes $\mathbf{x}_{t_i} \in S^2$. Therefore, if $\mathbf{x}_0$ is the initial direction of the axis and $g(t)$ describes the rigid body motion, then $g(t_i)\mathbf{x}_0 - \mathbf{x}_{t_i} = 0$, up to measurement error. One would like to model the trajectory $g(t)$, taking into account this information. The action functional (3.7) encodes one such model, yielding the curve $g(t)$ of minimal external torque (in the L^2 sense) that is consistent with the experiment. A natural choice for the parameter σ^2 is to set it equal to the variance of the measurement. Example simulations are shown in Figures 3.1 and 3.2. These were generated by the numerical algorithm discussed in Section 4.
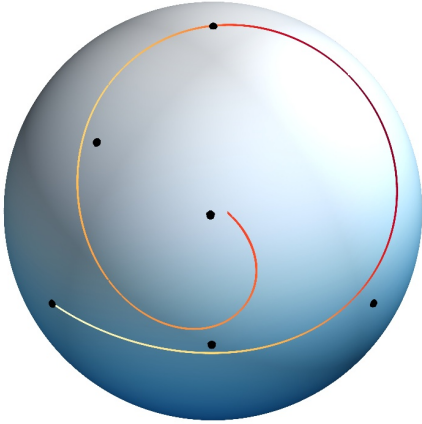

Fig. 3.1: Trajectory planning on the sphere through rigid rotations. The moment of inertia tensor was taken to be the identity matrix and the tolerance parameter was set to $\sigma = 0.025$. The given data points are represented as black dots with uniform time separation. The curve shown is $g(t)\mathbf{x}_0$ for the optimal curve $g(t)$, generated using the algorithm developed in Section 4. The coloring represents the local speed along the curve in $SO(3)$, that is, $\|\mathbf{\Omega}^0\|_{\mathfrak{so}(3)}$ (red is large, white is small).

If the observer is instead moving with a body fixed frame and measuring a space fixed direction, then $g(t_i)^{-1}\mathbf{x}_0 - \mathbf{x}_{t_i} = 0$, up to measurement error. In this case the inexact trajectory planning problem presents itself in the right-action, left-reduction form. Evidently, the formalism presented in this paper applies to any (sufficiently smooth) choice of Lagrangian. For example,

$$\ell(\mathbf{\Omega}^0, \mathbf{\Omega}^1) = \frac{1}{2}\mathbf{\Omega}^0 \cdot (\mathbb{I}\mathbf{\Omega}^1 + \mathbf{\Omega}^0 \times \mathbb{I}\mathbf{\Omega}^0)$$
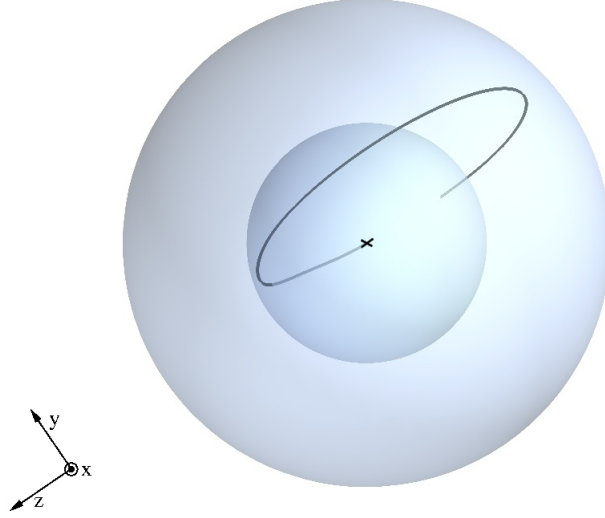
Fig. 3.2: Optimal curve $g(t)$ in the group $SO(3)$ of rigid rotations corresponding to the data points shown in Figure 3.1 and generated using the algorithm discussed in Section 4. The moment of inertia tensor was taken to be the identity matrix and the tolerance parameter was set to $\sigma = 0.025$. In this figure a given element of $SO(3)$ corresponds to a point on the radial line along the rotation axis, at a distance from the origin equal to the rotation angle. The radius of the inner sphere is π and the radius of the outer sphere is 2π. The center, marked by a cross, and the boundary of the outer sphere thus both represent the identity matrix.

leads to optimal curves $g(t)$ with minimal work done by external torques.

Remark 3.2. *We mentioned in Remark 3.1 that the minimization of (3.7) is related to a certain inverse problem given the stochastic evolution (3.3). Alternative stochastic models lead to forms of ℓ different from (3.6). For example,*

$$\mathrm{d}\boldsymbol{\Omega} = -\mathbb{I}^{-1}(\boldsymbol{\Omega} \times \mathbb{I}\boldsymbol{\Omega})\mathrm{d}t + \sigma_W \mathrm{d}\mathbf{W},$$

where $\mathrm{d}\mathbf{W} = (\mathrm{d}W^1, \mathrm{d}W^2, \mathrm{d}W^3)^T$ is a vector of independent Brownian motions, leads to

$$\ell(\boldsymbol{\Omega}^0, \boldsymbol{\Omega}^1) = \frac{1}{2} \left\| \boldsymbol{\Omega}^1 + \mathbb{I}^{-1}(\boldsymbol{\Omega}^0 \times \mathbb{I}\boldsymbol{\Omega}^0) \right\|^2$$

with $\|.\|$ being the Euclidean norm.

3.3   Quantum splines

In this section we give an overview of a quantum mechanical application presented in [BHM12], where more details can be found. We consider an $n + 1$ level quantum system with Hilbert space $\mathcal{H} = \mathbb{C}^{n+1}$. Quantum state space is the n-dimensional complex projective space $\mathbb{CP}^n$, which is defined as $\mathbb{C}^{n+1} - \{0\}$ modulo the equivalence relation $|\psi\rangle \sim \lambda|\psi\rangle$ for any complex number $\lambda \neq 0$. The standard geodesic distance is given by

$$d(\psi, \phi) = 2 \arccos \sqrt{\frac{\langle\psi|\phi\rangle\langle\phi|\psi\rangle}{\langle\psi|\psi\rangle\langle\phi|\phi\rangle}}.$$

Here we used Dirac notation to write $|\psi\rangle$ for a vector in $\mathbb{C}^{n+1}$ and $\langle\psi|$ for its Hermitian conjugate. The evolution of a quantum state is given by the *Schrödinger equation*

$$\frac{\mathrm{d}}{\mathrm{d}t}|\psi\rangle = -\mathrm{i}H|\psi\rangle,$$

where the Hamiltonian operator H is a Hermitian matrix. Equivalently one may express this as a differential equation on the unitary group $U(n+1)$ by

$$\dot{U} = -\mathrm{i}HU, \quad |\psi(t)\rangle = U(t)|\psi(0)\rangle. \tag{3.8}$$

One may assume without loss of generality that the anti-Hermitian matrix $\mathrm{i}H$ is of zero trace, therefore $U(t)$ is in the special unitary group $SU(n+1)$. This is due to the fact that the trace contributes a complex phase factor to the state evolution, which can be neglected in projective terms. Notice that $-\mathrm{i}H$ is skew-Hermitian and trace free and therefore lies in the Lie algebra $\mathfrak{su}(n+1)$.

Suppose a time dependent Hamiltonian $H(t)$ is controlled in an experiment whose goal it is to steer a system from initial state $|\psi_0\rangle$ through states $|\psi_{t_j}\rangle$ at times t_j, for $j = 1, \ldots, l$. One would like to achieve this trajectory with least change to the experimental apparatus. We formalize this requirement using an inner product on $\mathfrak{su}(n+1)$,

$$\gamma_e(A, B) = -2\operatorname{tr}(AB). \tag{3.9}$$

The cost we associate with a time dependent Hamiltonian $H(t)$ is $\int \frac{1}{2}\gamma_e(\dot{H}, \dot{H})\,\mathrm{d}t$. We note that γ_e extends to a bi-invariant Riemannian metric γ on $SU(n+1)$, which means in particular that $\operatorname{ad}^\dagger = -\operatorname{ad}$. The Lagrangian $\ell(H^0, H^1) = \frac{1}{2}\gamma_e(H^1, H^1)$ is therefore equal to the reduced Lagrangian (3.2) for the metric γ on $SU(n+1)$. The total cost functional takes the left-action, right-reduction form

$$S[U, \mathrm{i}H^0, \mathrm{i}H^1, M^0, M^1] = \int_0^{t_l} \operatorname{tr}(H^1 H^1) + \left\langle M^0, \dot{U}U^{-1} - \mathrm{i}H^0 \right\rangle + \left\langle M^1, \mathrm{i}\dot{H}^0 - \mathrm{i}H^1 \right\rangle \mathrm{d}t$$

$$+ \frac{1}{2\sigma^2}\sum_{i=1}^{l} d^2(U(t_i)\psi_0, \psi_{t_i}).$$

The tolerance parameter σ can be used to trade off quality of matching against a reduced cost associated with change in the Hamiltonian over time. For more details, see [BHM12].

3.4   Macromolecular configurations

Equilibrium configurations of macromolecular structures and of DNA in particular can be modelled using the classical theory of elastic rods. See [SH94, KC06] for examples of this approach. In this section we formulate an inexact trajectory planning problem in this context.

We start by describing how the configuration of an elastic rod can be described by a position curve $\mathbf{r}(s)$ in $\mathbb{R}^3$ and a curve $R(s)$ in the group of rigid rotations, $SO(3)$. Here, $s \in [0, 1]$ parameterizes the cross sections of the rod along its length, whereby for a given value of s the vector $\mathbf{r}(s)$ points to the center of mass of the respective cross section, as

seen in the lab frame. Let $\mathbf{e}_i(s)$, $i = 1, 2, 3$ be an orthonormal frame such that $\mathbf{e}_1(s)$ and $\mathbf{e}_2(s)$ point along the principal axes of the moment of inertia tensor of the cross section. We will refer to this frame as the *body-fixed* frame. $R(s)$ is the rotation that transforms the initial frame at $s = 0$ (which we assume to coincide with the lab frame) to the body-fixed frame at s. Therefore the configuration of a macromolecule can be described by a curve $g(s) = (R(s), \mathbf{r}(s))$ in the special Euclidean group, $SE(3)$, originating at the identity. The group multiplication rule of $SE(3)$ is

$$(R_1, \mathbf{r}_1)(R_2, \mathbf{r}_2) = (R_1 R_2, R_1 \mathbf{r}_2 + \mathbf{r}_1).$$

The Lie algebra $\mathfrak{se}(3)$ consists of elements $(\Omega, \mathbf{v})$ with $\Omega \in \mathfrak{so}(3)$ and $\mathbf{v} \in \mathbb{R}^3$. Applying the inverse of the hat map (3.4) to Ω we can represent $\mathfrak{se}(3)$ as $\mathbb{R}^6$. The ad operation becomes

$$\mathrm{ad}_{\boldsymbol{\xi}_1} \boldsymbol{\xi}_2 = \mathrm{ad}_{(\boldsymbol{\Omega}_1, \mathbf{v}_1)}(\boldsymbol{\Omega}_2, \mathbf{v}_2) = (\boldsymbol{\Omega}_1 \times \boldsymbol{\Omega}_2, \boldsymbol{\Omega}_1 \times \mathbf{v}_2 - \boldsymbol{\Omega}_2 \times \mathbf{v}_1).$$

If we identify the dual $\mathfrak{se}(3)^*$ with $\mathbb{R}^6$ using the standard dot product as duality pairing, then

$$\mathrm{ad}^*_{(\boldsymbol{\Omega}, \mathbf{v})}(\boldsymbol{\mu}, \mathbf{a}) = (-\boldsymbol{\Omega} \times \boldsymbol{\mu} - \mathbf{v} \times \mathbf{a}, -\boldsymbol{\Omega} \times \mathbf{a}). \tag{3.10}$$

The *body-fixed velocity* vector pertaining to a configuration $(R(s), \mathbf{r}(s))$ is defined as

$$\xi = g^{-1} \dot{g} = (R^{-1} \dot{R}, R^{-1} \dot{\mathbf{r}}) = (\Omega, R^{-1} \dot{\mathbf{r}})$$

where a superscript $\cdot$ means derivation with respect to s. In vector notation we can set $\mathbf{v} = R^{-1} \dot{\mathbf{r}}$ so that $\boldsymbol{\xi} = (\boldsymbol{\Omega}, \mathbf{v}) \in \mathbb{R}^6$.

A macromolecule that is experimentally constrained to assume a configuration with given final rotation and displacement $g(1) = (R(1), \mathbf{r}(1))$ will relax into an equilibrium state that minimizes potential energy with respect to all possible configurations respecting the constraint. For the case of DNA the authors of [KC06] propose to model this effect using the Lagrangian $L : TSE(3) \to \mathbb{R}$ whose left-reduced form $l : \mathfrak{se}(3) \to \mathbb{R}$ is given by

$$l(\boldsymbol{\xi}) = \frac{1}{2}(\boldsymbol{\xi} - \mathbf{z}) \cdot K(\boldsymbol{\xi} - \mathbf{z}).$$

The 6×6 matrix K is symmetric and positive definite and encodes the various stiffness properties. The double helix structure of DNA means that the equilibrium configuration for unconstrained end points retains a number n of rotations along its length. This is expressed by the vector

$$\mathbf{z} = \begin{pmatrix} 2\pi n \mathbf{e}_z \\ \mathbf{e}_z \end{pmatrix}, \tag{3.11}$$

where we assume without loss of generality that the equilibrium configuration for unconstrained end points is oriented along the spatial z-axis and has unit length.

In the case of constrained final rotation and displacement $g(1) = (R(1), \mathbf{r}(1))$ the equilibrium configuration minimizes the action functional $S = \int_0^1 l(\boldsymbol{\xi}) \, \mathrm{d}s$. Hamilton's principle $\delta S = 0$ leads to Euler–Poincaré equations

$$\dot{\boldsymbol{\xi}} = K^{-1} \mathrm{ad}^*_{\boldsymbol{\xi}} K(\boldsymbol{\xi} - \mathbf{z}) = \mathrm{ad}^{\dagger}_{\boldsymbol{\xi}}(\boldsymbol{\xi} - \mathbf{z}), \tag{3.12}$$

with ad^* operation given by (3.10) and the operation $\text{ad}^\dagger$ defined by $\text{ad}^\dagger_{\boldsymbol{\xi}_1} \boldsymbol{\xi}_2 := K^{-1} \text{ad}^*_{\boldsymbol{\xi}_1} K \boldsymbol{\xi}_2$. This equation of motion can be used to design second order Lagrangians to model non-equilibrium states of the DNA. For example, let us set

$$\ell(\boldsymbol{\xi}^0, \boldsymbol{\xi}^1) = \frac{1}{2}(\boldsymbol{\xi}^1 - \text{ad}^\dagger_{\boldsymbol{\xi}^0}(\boldsymbol{\xi}^0 - \mathbf{z})) \cdot K(\boldsymbol{\xi}^1 - \text{ad}^\dagger_{\boldsymbol{\xi}^0}(\boldsymbol{\xi}^0 - \mathbf{z}))$$
$$= \frac{1}{2} \left\| \boldsymbol{\xi}^1 - \text{ad}^\dagger_{\boldsymbol{\xi}^0}(\boldsymbol{\xi}^0 - \mathbf{z}) \right\|^2_K .$$

Suppose an experiment measures the position of the center of mass $\mathbf{r}(s_i)$ at a number of parameter values s_i $(i = 1, \ldots, l)$ as well as a body fixed direction, say $\mathbf{e}_3(s_i)$. The space of measurement outcomes is $Q = S^2 \times \mathbb{R}^3 \in \mathbb{R}^6$ with $SE(3)$ action given by

$$(R, \mathbf{r})(\widehat{\mathbf{x}}, \mathbf{y}) = (R\widehat{\mathbf{x}}, R\mathbf{y} + \mathbf{r}).$$

If the measurements yield the sequence $(\widehat{\mathbf{x}}_{s_i}, \mathbf{y}_{s_i})$ $(i = 1, \ldots, l)$ this suggests that up to measurement error the configuration $g(s) \in SE(3)$ satisfies

$$g(s_i)(\mathbf{e}_3(0), \mathbf{0}) = (\widehat{\mathbf{x}}_{s_i}, \mathbf{y}_{s_i}).$$

The task of modelling the configuration $g(s)$ can then be cast in the form of an inexact trajectory planning problem in the left-action, left-reduction form with cost functional

$$S[g, \boldsymbol{\xi}^0, \boldsymbol{\xi}^1, \boldsymbol{\mu}^0, \boldsymbol{\mu}^1] = \int_0^{t_l} \frac{1}{2} \left\| \boldsymbol{\xi}^1 - \text{ad}^\dagger_{\boldsymbol{\xi}^0}(\boldsymbol{\xi}^0 - \mathbf{z}) \right\|^2_K + \left\langle \boldsymbol{\mu}^0, g^{-1}\dot{g} - \boldsymbol{\xi}^0 \right\rangle + \left\langle \boldsymbol{\mu}^1, \dot{\boldsymbol{\xi}}^0 - \boldsymbol{\xi}^1 \right\rangle \, \mathrm{d}s$$
$$+ \frac{1}{2\sigma^2} \sum_{i=1}^{l} \left\| g(s_i)(\mathbf{e}_3(0), \mathbf{0}) - (\widehat{\mathbf{x}}_{s_i}, \mathbf{y}_{s_i}) \right\|^2 .$$

Remark 3.3. *Due to the anisotropy in velocity space, expressed by the vector* $\mathbf{z}$, *the Euler–Poincaré equation* (3.12) *is* not *the reduced geodesic equation for the curve* $g(t)$ *with respect to the metric defined by* K. *Consequently, the solution to the inexact trajectory matching problem is not a Riemannian cubic spline.*

4   Geometric discretization

The purpose of this section is to illustrate how the higher-order Hamilton–Pontryagin principle offers a direct route towards geometric numerical integrators. All one needs to do, in essence, is to provide a geometric discretization of the constraint $TR_{g^{-1}}\dot{g} - \xi^0 = 0$ and define a discrete Hamilton–Pontryagin principle accordingly. For first order variational problems this idea was introduced in [BRM09], building on the general theory of variational integrators (see [MW01] for an extensive review). We follow in the footsteps of [BRM09] to treat second order problems. Third and higher orders can be dealt with in a similar fashion.

Our main motivation for the development of a geometric integrator lies with the exact momentum behaviour exhibited by the discrete solution curves. Not only does the momentum conservation on open intervals (2.24) translate exactly into the discrete picture, but the

behavior at the nodes is given by discrete versions of the continuous time node equations (2.10)–(2.13). As a consequence, one can obtain discrete analogues of Theorem 2.2 and Corollary 2.3. This means that the numerical search for the optimal initial value of the momentum μ^0 can be restricted to a linear subspace of $\mathfrak{g}^*$ of the same dimension as the data manifold Q. As we shall see below, the variational nature of the integrator also means that the discrete flow map preserves the canonical symplectic form on $T^*(TG)$.

4.1   A geometric integrator

In discrete mechanics the time axis $[t_0, t_l]$ is replaced by a set of discrete time points $t_k = t_0 + kh$, $k = 0, \ldots, N$, where h is the step size and $t_l = t_0 + Nh$. We use integers N_i, $i = 1, \ldots, l$, as node indices $t_i = t_0 + N_i h$. For convenience we also define $N_0 := 0$. We will need a map $\tau : \mathfrak{g} \to G$ that approximates the Lie exponential and is an analytic diffeomorphism in a neighbourhood of 0 with $\tau(0) = e$ as well as $\tau(\xi)\tau(-\xi) = e$ for all $\eta \in \mathfrak{g}$. An example is the *Cayley transform*, which is a second-order approximation of the Lie exponential in quadratic matrix Lie groups. This includes the applications discussed in Section 3. The Cayley transform is defined as

$$\tau(\xi) = (e - \xi/2)^{-1}(e + \xi/2).$$

This and further examples are discussed in [BRM09].

Since τ is an approximate of the Lie exponential, a simple way of discretizing the constraint $TR_{g^{-1}}\dot{g} - \xi^0 = 0$ is to require that $g_{k+1} = \tau(h\xi_k^0)g_k$, where h is the size of a time step. Similarly one may translate $\dot{\xi}^0 = \xi^1$ to $\xi_{k+1}^0 = \xi_k^0 + h\xi_k^1$. With these considerations in mind we define a discretized version of the action functional (2.2) on discrete path space $\mathcal{C}_d$. Let

$$\mathcal{C}_d = \left\{ g_0, \xi_0^0, \xi_0^1, (g_k, \xi_k^0, \xi_k^1, \check{\mu}_k^0, \mu_k^1)_{k=1}^N \right\} = (G \times 2\mathfrak{g}) \times (G \times 2\mathfrak{g} \times 2\mathfrak{g}^*)^N,$$

then we define $S_d : \mathcal{C}_d \to \mathbb{R}$ as

$$S_d = h\left[\sum_{k=0}^{N-1} \ell(\xi_k^0, \xi_k^1) + \left\langle \check{\mu}_{k+1}^0, \frac{1}{h}\tau^{-1}(g_{k+1}g_k^{-1}) - \xi_k^0 \right\rangle + \left\langle \mu_{k+1}^1, \frac{1}{h}(\xi_{k+1}^0 - \xi_k^0) - \xi_k^1 \right\rangle \right]$$

$$+ \frac{1}{2\sigma^2} \sum_{i=1}^{l} d^2(g_{N_i}Q_0, Q_{t_i}). \tag{4.1}$$

In analogy to continuous time we assume that $g_0 = e$ and ξ_0^0 are given.

The discrete Euler–Lagrange equations follow from Hamilton's principle $\delta S_d = 0$. That is, they characterize paths $\gamma \in \mathcal{C}_d$, for which $\delta S_d := \delta \gamma(S) = 0$ for all variations $\delta \gamma \in T_\gamma \mathcal{C}_d$ with $\delta g_0 = 0$ and $\delta \xi_0^0 = 0$. In the process of computing δS_d we need to calculate $\delta \tau^{-1}(g_{k+1}g_k^{-1})$. For that purpose it is convenient to introduce the left-trivialized differential of τ at $\xi \in \mathfrak{g}$,

$$D\tau_\xi : \mathfrak{g} \to \mathfrak{g}, \quad \eta \mapsto \tau(\xi)^{-1} \left(T_\xi \tau(\eta) \right),$$

whose inverse we denote by $D\tau_\xi^{-1}$. By taking a derivative of $\tau(\xi)\tau(-\xi) = e$ one can show that [BRM09]

$$D\tau_\xi^{-1} = D\tau_{-\xi}^{-1} \circ \mathrm{Ad}_{\tau(\xi)}. \tag{4.2}$$

Denoting $\eta_k := (\delta g_k) g_k^{-1}$ we find

$$
\begin{aligned}
\delta\tau^{-1}(g_{k+1}g_k^{-1}) &= T_{\tau(h\xi_k^0)}\tau^{-1}(\eta_{k+1}\tau(h\xi_k^0)) - T_{\tau(h\xi_k^0)}\tau^{-1}(\tau(h\xi_k^0)\eta_k) \\
&= T_{\tau(h\xi_k^0)}\tau^{-1}(\tau(h\xi_k^0)\,\mathrm{Ad}_{\tau(-h\xi_k^0)}\,\eta_{k+1}) - T_{\tau(h\xi_k^0)}\tau^{-1}(\tau(h\xi_k^0)\eta_k) \\
&= D\tau_{h\xi_k^0}^{-1}(\mathrm{Ad}_{\tau(-h\xi_k^0)}\,\eta_{k+1)}) - D\tau_{h\xi_k^0}^{-1}(\eta_k) \\
&= D\tau_{-h\xi_k^0}^{-1}(\eta_{k+1}) - D\tau_{h\xi_k^0}^{-1}(\eta_k),
\end{aligned}
$$

where in the last equality we used (4.2). Introducing the quantities

$$
\mu_0^1 := \mu_1^1 + h\check{\mu}_1^0 - h\frac{\delta\ell}{\delta\xi_0^0}, \qquad \mu_0^0 := (D\tau_{h\xi_0^0}^{-1})^* \check{\mu}_1^0 \tag{4.3}
$$

and

$$
\mu_k^0 := (D\tau_{-h\xi_{k-1}^0}^{-1})^* \check{\mu}_k^0 \qquad (k = 1, \ldots, N), \tag{4.4}
$$

we obtain, after rearranging terms,

$$
\begin{aligned}
\delta S_d = h &\left[\sum_{k=1}^{N-1} \left\langle \frac{\delta\ell}{\delta\xi_k^0} - \check{\mu}_{k+1}^0 + \frac{1}{h}\mu_k^1 - \frac{1}{h}\mu_{k+1}^1, \delta\xi_k^0 \right\rangle + \left\langle \frac{\delta\ell}{\delta\xi_k^1} - \mu_{k+1}^1, \delta\xi_k^1 \right\rangle \right. \\
&\left. + \left\langle \frac{1}{h}\mu_k^0 - \frac{1}{h}(D\tau_{h\xi_k^0}^{-1})^*(D\tau_{-h\xi_k^0}^{-1})^*\mu_{k+1}^0 \right\rangle + \langle \delta\check{\mu}_{k+1}^0, \ldots \rangle + \langle \delta\mu_{k+1}^1, \ldots \rangle \right] \\
&+ h\left\langle \frac{\delta\ell}{\delta\xi_0^1} - \mu_1^1, \delta\xi_0^1 \right\rangle + h\langle \delta\check{\mu}_1, \ldots \rangle + h\langle \delta\mu_1^1, \ldots \rangle + \langle \mu_N^1, \delta\xi_N^0 \rangle - \langle \mu_0^1, \delta\xi_0^0 \rangle \\
&+ \langle \mu_N^0, \eta_N \rangle - \langle \mu_0^0, \eta_0 \rangle + \frac{1}{\sigma^2}\sum_{i=1}^{l} \langle d_{N_i} J^Q(\mathrm{d}_1 d_{N_i}), \eta_{N_i} \rangle,
\end{aligned}
$$

where we used abbreviations $d_{N_i} = d(g_{N_i}Q_0, Q_{t_i})$ and $\mathrm{d}_1 d_{N_i} := \mathrm{d}_1 d(g_{N_i}Q_0, Q_{t_i})$. The Euler–Lagrange equations are therefore composed of the following equations. The constraints

$$
g_{k+1} = \tau(h\xi_k^0)g_k, \qquad \xi_{k+1}^0 = \xi_k^0 + h\xi_k^1 \qquad (k = 0, \ldots, N-1), \tag{4.5}
$$

the discrete equations for the Legendre–Ostrogradsky momenta

$$
\mu_{k+1}^1 = \mu_k^1 - h(D\tau_{-h\xi_k^0}^{-1})^*\mu_{k+1}^0 + h\frac{\delta\ell}{\delta\xi_k^0} \qquad (k = 1, \ldots, N-1) \tag{4.6}
$$

$$
\frac{\delta\ell}{\delta\xi_k^1} - \mu_{k+1}^1 = 0 \qquad (k = 0, \ldots, N-1), \tag{4.7}
$$

the discrete version of the Euler–Poincaré equation for interior indices $k \neq N_i$ $(i = 1, \ldots, l)$

$$
\mu_{k+1}^0 = (D\tau_{-h\xi_k^0}^{-1})^*(D\tau_{h\xi_k^0})^*\mu_k^0 \tag{4.8}
$$

and for node indices $k = N_i$ $(i = 1, \ldots, l-1)$

$$
\mu_{k+1}^0 = (D\tau_{-h\xi_k^0}^{-1})^*(D\tau_{h\xi_k^0})^* \left(\mu_k^0 + \frac{d_k}{\sigma^2} J^Q(\mathrm{d}_1 d_k) \right). \tag{4.9}
$$

Finally,

$$\mu_N^0 + \frac{d_N}{\sigma^2} J^Q(\mathrm{d}_1 d_N) = 0, \tag{4.10}$$

$$\mu_N^1 = 0. \tag{4.11}$$

A solution $\gamma \in \mathcal{C}_d$ to (4.5)–(4.9) is said to have *initial conditions* $(g_0, \xi_0^0, \mu_0^0, \mu_0^1) \in G \times \mathfrak{g} \times 2\mathfrak{g}^* \cong T^*(TG)$, using the definitions (4.3). If the Lagrangian ℓ is hyperregular, then (4.7) can be solved for ξ_k^1. This means that ξ_k^1 can be eliminated from equations (4.5)–(4.9), which can subsequently be integrated for given initial conditions. If in addition equations (4.10) and (4.11) are satisfied, then γ is a critical point of the action functional S_d.

Remark 4.1. *In Remark 2.1 we mentioned that besides the above left-action, right-reduction case, three other cases were available to be considered. In a similar manner to what we observed in that remark the right-action, right-reduction case introduces changes to equations (4.9) and (4.10). These become*

$$\mu_{k+1}^0 = (D\tau_{-h\xi_k^0}^{-1})^*(D\tau_{h\xi_k^0})^*\left(\mu_k^0 - \frac{d_k}{\sigma^2}\,\mathrm{Ad}^*_{g_k^{-1}} J^Q(\mathrm{d}_1 d_k)\right) \quad and$$

$$\mu_N^0 - \frac{d_N}{\sigma^2}\,\mathrm{Ad}^*_{g_N^{-1}} J^Q(\mathrm{d}_1 d_N) = 0,$$

respectively. The remaining two cases (left-action, left-reduction; and right-action, left-reduction) are equivalent to the first two in just the same way as explained in Remark 2.1.

4.2   Geometric properties

As in Section 2.3 the interior equations are conveniently analyzed by omitting the mismatch penalty term (4.1) from the action functional, so that

$$\mathcal{J}_d = h\left[\sum_{k=0}^{N-1} \ell(\xi_k^0, \xi_k^1) + \left\langle \check{\mu}_{k+1}^0, \frac{1}{h}\tau^{-1}(g_{k+1}g_k^{-1}) - \xi_k^0 \right\rangle + \left\langle \mu_{k+1}^1, \frac{1}{h}(\xi_{k+1}^0 - \xi_k^0) - \xi_k^1 \right\rangle\right]$$

The arguments that surround equation (2.23) and show symplecticity of the continuous time flow can then be applied in a straightforward manner to the discrete case. Indeed, interior equations (4.5)–(4.8) define a flow map $F_d : G \times \mathfrak{g} \times 2\mathfrak{g}^* \to G \times \mathfrak{g} \times 2\mathfrak{g}^*$, which integrates a solution γ for given initial conditions. That is, $(F_d)^k(g_0, \xi_0^0, \mu_0^0, \mu_0^1) = (g_k, \xi_k^0, \mu_k^0, \mu_k^1)$ or more succinctly $(F_d)^k(\gamma_0) = \gamma_k$. We restrict $\mathcal{J}_d$ to solutions of (4.5)–(4.8) and express it as a function $\mathcal{J}_{d,\text{initial}} : T^*(TG) \to \mathbb{R}$ of initial conditions $\gamma_0 \in T^*(TG)$. This means that if $\gamma \in \mathcal{C}_d$ is the solution obtained by integrating γ_0 then $J_{d,\text{initial}}(\gamma_0) = \mathcal{J}_d(\gamma)$. It follows that

$$\mathrm{d}\mathcal{J}_{d,\text{initial}}(\delta\gamma_0) = \left\langle \mu_N^1, \delta\xi_N^0 \right\rangle - \left\langle \mu_0^1, \delta\xi_0^0 \right\rangle + \left\langle \mu_N^0, \eta_N \right\rangle - \left\langle \mu_0^0, \eta_0 \right\rangle = (((F_d)^k)^*\theta - \theta)(\delta\gamma_0),$$

where θ is the canonical one-form on $T^*(TG)$. Taking an exterior derivative shows that the canonical symplectic form $\omega = \mathrm{d}\theta$ is preserved by F_d. Hence the discrete flow map F_d is symplectic.

Similarly the observations given in Section 2.4 can be translated to the discrete picture. Indeed we pointed out in the paragraph of equation (2.25) how to obtain the conservation of the momentum $J_k = \mathrm{Ad}^*_g\, \mu^0$ from a variational perspective. These arguments can be applied to the discrete variational principle to show that $\mathrm{Ad}^*_{g_{k+1}} \mu^0_{k+1} = \mathrm{Ad}^*_{g_k} \mu^0_k$ for interior indices $k \neq N_i$. Alternatively, a manipulation of equation (4.8) using (4.2) shows that

$$\mu^0_{k+1} = (D\tau^{-1}_{-h\xi^0_k})^*(D\tau_{h\xi^0_k})^*\mu^0_k = \mathrm{Ad}^*_{\tau(-h\xi^0_k)} \mu^0_k = \mathrm{Ad}^*_{g_k g^{-1}_{k+1}} \mu^0_k = \mathrm{Ad}^*_{g^{-1}_{k+1}} \mathrm{Ad}^*_{g_k} \mu^0_k,$$

and therefore $\mathrm{Ad}^*_{g_{k+1}} \mu^0_{k+1} = \mathrm{Ad}^*_{g_k} \mu^0_k$.

Remark 4.2. *The symplectic flow equations on open intervals can equivalently be discretized in the purely Lagrangian picture by following [CJdD12]. One chooses a suitable discrete Lagrangian $L_d : G \times G \times G \to \mathbb{R}$ and applies variational calculus to the discrete action sum*

$$S_d = \sum_{k=0}^{N-2} L_d(g_k, g_{k+1}, g_{k+2}).$$

Let us define $\xi : G \times G \to \mathfrak{g}$ by $\xi(g_1, g_2) = h^{-1}\tau^{-1}\left(g_2 g^{-1}_1\right)$ and set

$$L_d(g_k, g_{k+1}, g_{k+2}) := h\ell(\xi(g_k, g_{k+1}), h^{-1}(\xi(g_{k+1}, g_{k+2}) - \xi(g_k, g_{k+1}))).$$

Boundary conditions being equal, the resulting optimal curve $(g_0, \dots, g_N)$ is the same as the one we obtained from the discrete Hamilton–Pontryagin principle. The Hamilton–Pontryagin principle has an advantage in situations where more sophisticated discretizations of the constraint $TR_{g^{-1}}\dot{g} = \xi^0$ are chosen. For example, in the preliminary study [BHM11] Runge–Kutta–Munthe–Kaas methods were used to introduce a class of such integrators. Those integrators can still be understood in the purely Lagrangian framework, however the definition of the corresponding function $L_d(g_k, g_{k+1}, g_{k+2})$ is implicit in that evaluating it requires solving a variational problem. The Hamilton–Pontryagin approach circumvents this difficulty by building the discretization of the constraint into the variational principle from the outset.

The node equation (4.9) reflects in a geometrically consistent way the jump discontinuities of $\mathrm{Ad}^*_g\, \mu^0$ given in (2.29). Indeed, (4.9) says that

$$J_{k+1} = J_k + \frac{d_k}{\sigma^2} \mathrm{Ad}^*_{g_k} J^Q(\mathrm{d}_1 d_k)$$

when $k = N_i$. Moreover, the final time conditions (4.10) and (4.11) are exact analogues of (2.12) and (2.13). See Figure 4.1 for an example. Putting everything together leads to discrete versions of Theorem 2.2 and Corollary 2.3. The proofs are analogous to the continuous time case.

Theorem 4.3. *For $k = 0, \dots, N$*

$$\mu^0_k = -\frac{1}{\sigma^2} \mathrm{Ad}^*_{g^{-1}_k} \left(\sum_{i=1}^{l} 1_{k \leq N_i} d_{N_i} \mathrm{Ad}^*_{g_{N_i}} J^Q(\mathrm{d}_1 d_{N_i}) \right).$$

Corollary 4.4. *For $k = 0, \dots, N$*

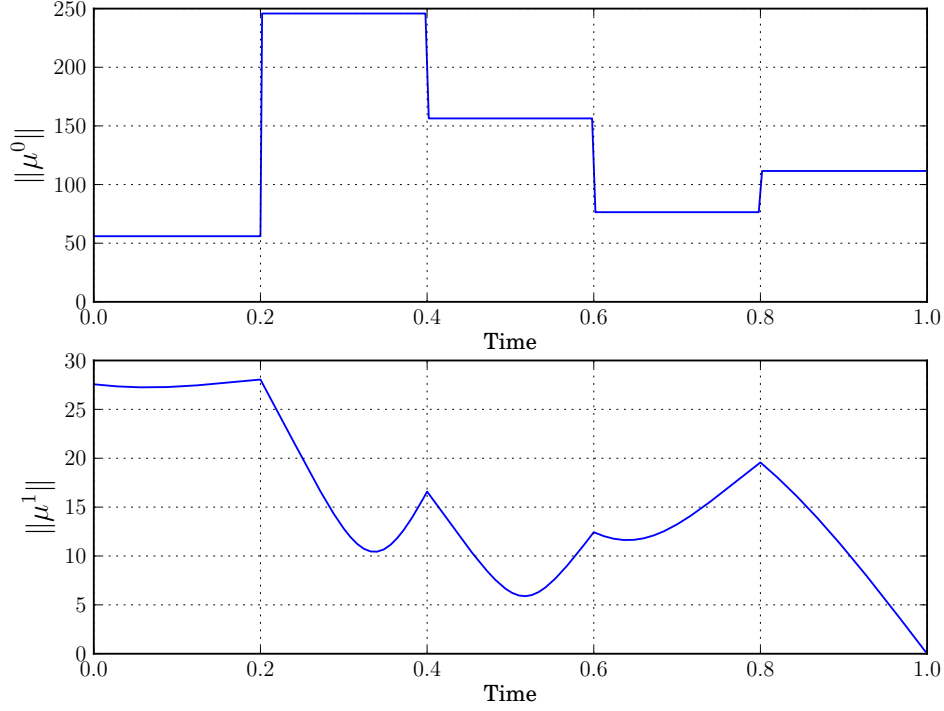$$\langle \mu^0_k, \rho \rangle = 0 \qquad \text{for all } \rho \in \mathfrak{g}_{g_k Q_0}.$$

Fig. 4.1: Momentum norms. For the interpolating discrete cubic of Figure 3.2, the plot shows the norms of the momenta μ_k^0 and μ_k^1. The norm of μ_k displays the momentum discontinuities at node indices as well as exact conservation at interior indices, in accordance with (4.8) and (4.9). The norm of μ_k^1 demonstrates continuity, as found in (4.6). Both curves respect terminal conditions (4.10) and (4.11).

4.3   Practical remarks

The integrator derived above provides a way of finding a numerical solution to the inexact trajectory planning problem.

The discrete equations of motion (4.5)–(4.9) can be employed to express the action functional $\mathcal{J}_d$ as a function $\mathcal{J}_{d,\text{initial}} : T^*(TG) \to \mathbb{R}$ of initial conditions $\gamma_0 = (g_0, \xi_0^0, \mu_0^0, \mu_0^1) \in T^*(TG) \cong G \times \mathfrak{g} \times 2\mathfrak{g}^*$. The minimizer in the space of initial conditions can then be determined by a gradient descent method. Since $g_0 = e$ and ξ_0^0 are given, the minimization is in effect only over the variables (μ_0^0, μ_0^1). By Corollary 4.4 the optimal μ_0^0 lies in the subspace of $\mathfrak{g}^*$ that annihilates $\mathfrak{g}_{Q_0}$, to which the search can therefore be restricted. On the other hand one still needs to consider all of $\mathfrak{g}^*$ for the optimization of μ_0^1.

The gradient of $\mathcal{J}_{d,\text{initial}}$ can be estimated via finite-difference methods. However, this requires the repeated forward integration of (4.5)–(4.9). The number of such integrations increases with the number of dimensions of the Lie group G, and for higher dimensional systems this quickly becomes unfeasible. Such difficulties can be circumvented using the standard method of *adjoint equations*, which can be easily implemented for the geometric discretization presented here (a detailed derivation is provided in the Appendix). Then the

exact gradient is obtained at the cost of integrating twice (once forward and once backward) a system of equations of the same complexity as the forward equations.

The significance of the exact preservation of final time constraints (2.12) and (2.13) in the form of (4.10), (4.11) is to provide verification that a (local) minimum has been found. When the tolerance is tight (σ is small) and in the absence of a good initial guess it may occur that the algorithm tends to a local minimum rather than the global one. Suitable initial guesses can be computed using a homotopy strategy. This means a step-by-step reduction of σ, where the optimum at one value of σ is taken as the initial guess at the next smaller value.

We used this algorithm to generate the figures in this paper. For simulations of quantum splines (see Section 3.3) using the same methods we refer to [BHM12].

5   Discussion and Outlook

Discussion.   This paper has discussed a type of inexact trajectory planning problem whose optimal curves are required to pass near a sequence of fixed target positions at designated times within a certain tolerance. In Section 2 a new derivation of the Euler–Lagrange equations for this type of problem was obtained by applying reduction by symmetry to a higher–order Hamilton–Pontryagin principle. This approach provided a new geometric interpretation of the previously known node equations in terms of Legendre–Ostrogradsky momenta. The highest order momentum was seen to undergo discontinuous jumps at the node times as a consequence of a partially broken Lie group symmetry. This was the content of Theorem 2.2 and Corollary 2.3. In Section 3 several applications of the theory were discussed, which summoned the inexact trajectory planning problem both from a control theoretic viewpoint (quantum splines) as well as in the context of a type of inverse problem (rigid body splines, macromolecular configurations). Finally, Section 4 was concerned with the numerical approach to solving the problem at hand. The reduced Hamilton–Pontryagin principle was taken as the starting point and a geometric discretization of the Euler–Lagrange equations was obtained, which led to exact momentum behavior and discrete versions of both Theorem 2.2 and Corollary 2.3. This meant in particular that the search for the optimal initial value of the highest order momentum could be restricted to a subspace of the dual of the Lie algebra of the Lie group G, whose action describes the motion.

Outlook.   This work invites further development in several directions. First, one can show that the discrete flow map derived in Section 4 is accurate only to first order in step size h. The development of geometric methods with a higher degree of accuracy would be desirable. The class of integrators presented in the preliminary study [BHM11] may prove useful in this regard. An alternative possibility is the Lagrangian approach of [CJdD12] together with suitably exact approximations of the Lagrangian function. Second, the final step in the numerical optimization of the cost functional was based on a shooting method with gradient descent in the space of initial conditions. A comparison with more sophisticated methods of nonlinear programming would be a useful guide for further development. Third, the example of the molecular strand (Section 3.4) was solved as a problem of statics. Adding the consid-

eration of dynamics of the strand brings one into the realm of so-called G-Strands [HIP12], for which the inexact trajectory planning problem may prove interesting. Finally, for applications in computational anatomy one must deal with infinite dimensional diffeomorphism groups. Extending the framework of this paper to infinite dimensions is therefore another challenge that lies ahead.

A   Gradient calculation via adjoint equations

In order to implement an efficient descent method for $\mathcal{J}_{d,\text{initial}}$ it is useful to have an expression for its gradient. One may obtain a gradient estimate using finite difference methods. One drawback lies with the inaccuracies inherent in the estimation. Moreover, if the dimension of the Lie algebra $\mathfrak{g}$ is large such estimations quickly become computationally costly. Both of these drawbacks can be circumvented in our case by the use of adjoint equations. In this way we obtain an exact expression for the gradient of $\mathcal{J}_{d,\text{initial}}$, in a computationally efficient way. We will derive the system of adjoint equations now. For simplicity we treat only the special case where $\ell = \frac{1}{2} \left\| \xi^1 \right\|_\gamma^2$, where γ denotes an inner product on $\mathfrak{g}$ and $\|.\|$ the corresponding norm. Moreover we will only treat the left-action, right-reduction case, however the others can be obtained in the same way. We recall from (4.5)–(4.9) the equations of motion

$$g_{k+1} = \tau(h\xi_k^0)g_k, \quad \xi_{k+1}^0 = \xi_k^0 + h(\mu_{k+1}^1)^\sharp \tag{A.1}$$

$$\mu_{k+1}^1 = \mu_k^1 - h(D\tau_{-h\xi_k^0})^* \mu_{k+1}^0, \tag{A.2}$$

$$\mu_{k+1}^0 = (D\tau_{-h\xi_k^0}^{-1})^* (D\tau_{h\xi_k^0})^* \left(\mu_k^0 + \Delta_k(g_k Q_0) \right), \tag{A.3}$$

where we introduce functions $\Delta_k : Q \to \mathfrak{g}^*$ for $k = 0, \dots, N$ defined as

$$\Delta_{N_i}(q) := \frac{d_{N_i}}{\sigma^2} J^Q(\mathrm{d}_1 d(g_{N_i} Q_0, Q_{t_l})$$

when $k \in \{N_1, \dots, N_l\}$ and $\Delta_k = 0$ otherwise.

Let us define an augmented functional $\mathcal{G}$, in which these equations are paired with Lagrange multipliers. These Lagrange multipliers will be denoted $(P_k^0, P_k^1, V_k^0, V_k^1) \in 2\mathfrak{g}^* \times 2\mathfrak{g}$ for $k = 1, \dots, N$. Let us introduce the shorthand notation x representing the discrete path $(g_k, \xi_k^0, \mu_k^0, \mu_k^1)_{k=0}^N$ and λ representing the ensemble of Lagrange multiplier $(P_k^0, P_k^1, V_k^0, V_k^1)_{k=1}^N$. The augmented functional $\mathcal{G}$ is given by

$$\mathcal{G}(x,\lambda) = h\left[\sum_{k=0}^{N-1} \frac{1}{2} \left\| (\mu_{k+1}^1)^\sharp \right\|_\gamma^2 + \left\langle P_{k+1}^0, \tau^{-1}(g_{k+1}g_k^{-1}) - h\xi_k^0 \right\rangle \right.$$

$$+ \left\langle P_{k+1}^1, \xi_{k+1}^0 - \xi_k^0 - h(\mu_{k+1}^1)^\sharp \right\rangle + \left\langle \mu_{k+1}^1 - \mu_k^1 + h(D\tau_{-h\xi_k^0})^* \mu_{k+1}^0, V_{k+1}^0 \right\rangle$$

$$+ \left\langle (D\tau_{-h\xi_k^0})^* \mu_{k+1}^0 - (D\tau_{h\xi_k^0})^* (\mu_k^0 + \Delta_k(g_k Q_0)), V_{k+1}^1 \right\rangle \right] + \frac{1}{2\sigma^2} \sum_{i=1}^{l} d^2(g_{N_i} Q_0, Q_{t_i}).$$

No constraints are assumed here, apart from the prescribed initial velocity ξ_0^0 and $g_0 = e$. It is clear that for any choice of Lagrange multipliers λ we have $\mathcal{G}(x, \lambda) = \mathcal{J}_{d,\text{initial}}(\mu_0^0, \mu_0^1)$, as long as x satisfies (A.1)–(A.3) for given initial values μ_0^0, μ_0^1. A tedious, but straightforward calculation shows that

$$\delta\mathcal{G}(x, \lambda) = -h \left\langle \delta\mu_0^1, V_1^0 \right\rangle - h \left\langle \delta\mu_0^0, D\tau_{h\xi_0^0} V_1^1 \right\rangle, \tag{A.4}$$

if x satisfies (A.1)–(A.3) *and* λ is a solution of the *adjoint equations*. We describe these now. We introduce functions $K^\pm : 2\mathfrak{g} \times \mathfrak{g}^* \to \mathfrak{g}^*$ by the defining relation

$$\left\langle K_{\xi,\mu}^\pm V, \rho \right\rangle = \left. \frac{\mathrm{d}}{\mathrm{d}\varepsilon} \right|_{\varepsilon=0} \left\langle (D\tau_{\pm h(\xi + \varepsilon\rho)})^* \mu, V \right\rangle, \quad \text{for all } \xi, V, \rho \in \mathfrak{g}, \mu \in \mathfrak{g}^*.$$

Moreover, for $k = 0, \dots, N$ we define functions $\mathcal{A}_k : Q \times \mathfrak{g} \to \mathfrak{g}^*$ by

$$\left\langle \mathcal{A}_k(q, \rho), \eta \right\rangle = \left. \frac{\mathrm{d}}{\mathrm{d}\varepsilon} \right|_{\varepsilon=0} \left\langle \Delta_k(\exp(\varepsilon\eta)q), \rho \right\rangle,$$

for all $q \in Q$ and $\rho, \eta \in \mathfrak{g}$. The adjoint equations consist of conditions at the final time point,

$$P_N^0 = -h^{-1}(D\tau_{-h\xi_{N-1}^0})^* \Delta_N(g_N Q_0), \quad P_N^1 = 0, \tag{A.5}$$

$$V_N^0 = -(\mu_N^1)^\sharp, \quad V_N^1 = -hV_N^0, \tag{A.6}$$

and the following equations for $k = 1, \dots, N-1$,

$$P_k^0 = (D\tau_{-h\xi_{k-1}^0})^* \left[(D\tau_{h\xi_k^0}^{-1})^* P_{k+1}^0 + \mathcal{A}_k(g_k Q_0, D\tau_{h\xi_k^0} V_{k+1}^1) - h^{-1}\Delta_k(g_k Q_0) \right] \tag{A.7}$$

$$P_k^1 = P_{k+1}^1 + hP_{k+1}^0 - hK_{\xi_k^0, \mu_{k+1}^0}^- V_{k+1}^0 - K_{\xi_k^0, \mu_{k+1}^0}^- V_{k+1}^1 + K_{\xi_k^0, \mu_k^0 + \Delta_k(g_k Q_0)}^+ V_{k+1}^1 \tag{A.8}$$

$$V_k^0 = V_{k+1}^0 - (\mu_k^1)^\sharp + h(P_k^1)^\sharp, \tag{A.9}$$

$$V_k^1 = -hV_k^0 + D\tau_{-h\xi_{k-1}^0}^{-1} D\tau_{h\xi_k^0} V_{k+1}^1. \tag{A.10}$$

These equations are posed backwards. That is, solving the adjoint equations entails initialising the Lagrange multipliers at time point N according to (A.5)–(A.6) and then iterating backwards from $k = N$ to $k = 1$ using (A.7)–(A.10).

We now obtain an expression for the gradient of $\mathcal{J}_{d,\text{initial}}$ from (A.4). Indeed, let $(\mu_0^0(\varepsilon), \mu_0^1(\varepsilon))$ be a variation of initial conditions (μ_0^0, μ_0^1), and let $x(\varepsilon)$ be the corresponding set of solutions to (A.1)–(A.3). Let λ be a solution to the adjoint equations (A.7)–(A.10) for $x = x(0)$, then

$$\delta\mathcal{J}_{d,\text{initial}} = \left. \frac{\mathrm{d}}{\mathrm{d}\varepsilon} \right|_{\varepsilon=0} \mathcal{J}_{d,\text{initial}}(\mu_0^0(\varepsilon), \mu_0^1(\varepsilon)) = \left. \frac{\mathrm{d}}{\mathrm{d}\varepsilon} \right|_{\varepsilon=0} \mathcal{G}(x(\varepsilon), \lambda)$$

$$= -h \left\langle \delta\mu_0^1, V_1^0 \right\rangle - h \left\langle \delta\mu_0^0, D\tau_{h\xi_0^0} V_1^1 \right\rangle.$$

From this we can read off the gradient,

$$\frac{\delta\mathcal{J}_{d,\text{initial}}}{\delta\mu_0^0} = -hD\tau_{h\xi_0^0} V_1^1, \qquad \frac{\delta\mathcal{J}_{d,\text{initial}}}{\delta\mu_0^1} = -hV_1^0.$$

References

[BHM11]   C. L. Burnett, D. D. Holm, and D. M. Meier. Geometric integrators for higher-order mechanics on Lie groups, December 2011. Available at http://arxiv.org/abs/1112.6037.

[BHM12]   D. C. Brody, D. D. Holm, and D. M. Meier. Quantum splines. *Phys. Rev. Lett.*, 109:100501, 2012.

[BRM09]   N. Bou-Rabee and J. E. Marsden. Hamilton Pontryagin integrators on Lie groups part i: Introduction and structure-preserving properties. *Found. Comput. Math.*, 9:197–219, March 2009.

[CdD11]   L. Colombo and D. M. de Diego. On the geometry of higher-order variational problems on Lie groups, April 2011. Preprint available at http://arxiv.org/abs/1104.3221.

[CJdD12]   L. Colombo, F. Jimenez, and D. M. de Diego. Discrete second-order Euler-Poincaré equations: Applications to optimal control. *International Journal of Geometric Methods in Modern Physics*, 09(04):1250037 – 1250057, 2012.

[CMR01]   H. Cendra, J.E. Marsden, and T.S. Ratiu. Lagrangian Reduction by Stages. *Memoirs of the Amer. Math. Soc.*, 152(722):1–117, 2001.

[CS95]   P. E. Crouch and F. Silva Leite. The dynamic interpolation problem: On Riemannian manifolds, Lie groups, and symmetric spaces. *Journal of Dynamical and Control Systems*, 1(2):177–202, 1995.

[CSC95]   M. Camarinha, F. Silva Leite, and P. E. Crouch. Splines of class C^k on non-Euclidean spaces. *IMA Journal of Mathematical Control & Information*, 12:399–410, 1995.

[CSC01]   M. Camarinha, F. Silva Leite, and P. Crouch. On the geometry of Riemannian cubic polynomials. *Differential Geometry and its Applications*, 15(2):107–135, 2001.

[dLR85]   M. de León and P. R. Rodrigues. Formalismes Hamiltonien symplectique sur les fibrés tangents d'ordre supérieur. *C. R. Acad. Sc. Paris*, 301:455–458, 1985.

[GBHM+12]   F. Gay-Balmaz, D. D. Holm, D. M. Meier, T. S. Ratiu, and F.-X. Vialard. Invariant higher-order variational problems. *Commun. Math. Phys.*, 309:413–458, 2012.

[GBHR11]   F. Gay-Balmaz, D. D. Holm, and T. S. Ratiu. Higher order Lagrange-Poincaré and Hamilton-Poincaré reductions. *Bulletin of the Brazilian Mathematical Society*, 42(4):579–606, 2011.

[GGP02]   R. Giambò, F. Giannoni, and P. Piccione. An analytical theory for Riemannian cubic polynomials. *IMA Journal of Mathematical Control & Information*, 19:445–460, 2002.

[HB04]   I. I. Hussein and A. M. Bloch. Dynamic interpolation on Riemannian manifolds: an application to interferometric imaging. In *Proceedings of the 2004 American Control Conference*, volume 1, pages 685 –690 vol.1, 2004.

[HIP12]   D. D. Holm, R. I. Ivanov, and J. R. Percival. *G*-strands. *Journal of Nonlinear Science*, 22(4):517–551, 2012.

[Hol11]   D. D. Holm. *Geometric Mechanics, Part II: Rotating, Translating and Rolling.* Imperial College Press, 2nd edition, 2011.

[KC06]   J. S. Kim and G. S. Chirikjian. Conformational analysis of stiff chiral polymers with end-constraints. *Molecular Simulation*, 32(14):1139–1154, 2006.

[MR03]   J. E. Marsden and T. S. Ratiu. *Introduction to Mechanics and Symmetry*, volume 17 of *Texts in Applied Mathematics*. Springer, New York, second edition, 2003.

[MTY02]   M. I. Miller, A. Trouvé, and L. Younes. On the metrics and Euler-Lagrange Equations of Computational Anatomy. *Annual Review of Biomedical Engineering*, 4:375–405, 2002.

[MW01]   J. E. Marsden and M. West. Discrete mechanics and variational integrators. *Acta Numerica*, 10:357–514, 2001.

[NHP89]   L. Noakes, G. Heinzinger, and B. Paden. Cubic Splines on Curved Spaces. *IMA Journal of Mathematical Control & Information*, 6:465–473, 1989.

[Noa04]   L. Noakes. Non-null Lie quadratics in E3. *J. Math. Phys.*, 45:4334–4351, 2004.

[Noa06]   L. Noakes. Duality and Riemannian cubics. *Adv. Comput. Math.*, 25:195–209, 2006.

[Øks03]   B. Øksendahl. *Stochastic Differential Equations: An Introduction with Application.* Universitext. Springer, 6th edition, 2003.

[PMRR11]   P. D. Prieto-Martínez and N. Román-Roy. Lagrangian-Hamiltonian unified formalism for autonomous higher-order dynamical systems. *J. Phys. A: Math. Theor.*, 44:385203, 2011.

[PR97]   F. C. Park and B. Ravani. Smooth Invariant Interpolation of Rotations. *ACM Transactions on Graphics*, 16:277–295, 1997.

[SH94]   Y. Shi and J. E. Hearst. The Kirchhoff elastic rod, the nonlinear Schrödinger equation, and DNA supercoiling. *The Journal of Chemical Physics*, 101(6):5186–5200, 1994.

[VT12]   F.-X. Vialard and A. Trouvé. Shape Splines and Stochastic Shape Evolutions: A Second Order Point of View. *Quart. Appl. Math.*, 70:219–251, 2012.

[YM06]   H. Yoshimura and J. E. Marsden. Dirac structures in Lagrangian mechanics part II: Variational structures. *Journal of Geometry and Physics*, 57(1):209–250, 2006.

[You10]   L. Younes. *Shapes and Diffeomorphisms*. Springer, 2010.

[ZKC98]   M. Zefran, V. Kumar, and C.B. Croke. On the generation of smooth three-dimensional rigid body motions. *IEEE Transactions on Robotics and Automation*, 14(4):576–589, 1998.